UNIVERSITY OF SOUTHAMPTON

STIMULUS EQUIVALENCE AND NAMING

Thomas David William Randell

Thesis submitted for Doctor of Philosophy

Department of Psychology
Faculty of Social Sciences

January 2000

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

PSYCHOLOGY

Doctor of Philosophy

STIMULUS EQUIVALENCE AND NAMING

by Thomas David William Randell

The functionality of verbal behaviour in stimulus equivalence was demonstrated by training verbally able adults with different combinations of easily nameable, yet formally unrelated, pictorial stimuli. Study One indicated that participants who were trained with combinations of pictures whose names rhymed with each other demonstrated the formation and generalisation of equivalence classes more readily than participants who were trained with non-rhyming combinations of the same stimuli. Studies Two and Three provided within-participant confirmations of this finding, and further indicated that previously established contextual control of baseline relations may be superseded by verbal control during testing without reinforcement. That verbal control and contextual cues may both provide a basis for the formation of generalised classes was also indicated. Study Four investigated the formation of contextually controlled equivalence classes using a think-aloud procedure, and additionally compared the performance of participants who were required to think-aloud during experimentation with that of participants who were not required so to do. The results indicated that use of such procedures may disrupt the formation of contextually controlled equivalence classes. Study Five demonstrated the emergence and generalisation of stimulus classes on the basis of verbal control in the absence of reinforcement baseline training. Overall, the findings indicated that visual stimuli are named, that the phonological properties of those names can influence equivalence class formation and generalisation, and that the emergence of untrained behaviour may, under certain circumstances, be verbally controlled.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I have been lucky enough to have received a great deal of support during the completion of this thesis. With this in mind, I wish firstly to thank Bob Remington, not only for his first rate supervision, but also for the friendship without which this thesis could not have been completed. I would also like to thank Derek Blackman for setting the occasion for my initial appreciation of radical behaviourism, and for the subsequent years of encouragement and judicious reinforcement. Thanks also to Bill Dube for his practical advice, musical anecdotes, and the free provision of excellent experimental software. To Maree Hunt, for her companionship and advice, and for making even proof reading enjoyable, what can I say? That was neat! My thanks also to Tony Roberts and Bryan Newman for their technical assistance during my research, and to all those who volunteered to participate in that research, my gratitude. To my good friends Ben Townsend and Faraz Rizvi, much respect guys--keep it right! To Anna-Maria Raffa, well, I hope I do still know you when I'm sixty-five. Lastly, for a lifetime of unflagging love and provision, and for the creation of almost all the experimental stimuli used in my research,

This thesis is dedicated to Jenny and Bill Randell, my parents

1. GENERAL INTRODUCTION

1.1 STRUCTURE OF THE THESIS

This thesis presents a programme of research, the aim of which was to assay the role of verbal behaviour in the formation and generalisation of equivalence classes. Chapter One describes the emergence of stimulus equivalence as an account of untrained behaviour. Chapter Two documents the empirical evidence of that theory's generality and provides an overview of two other major accounts of untrained behaviour. Chapter Three reviews methodological and procedural considerations relevant to equivalence research. Chapters Four to Eight present the research that forms the empirical basis of the thesis, and Chapter Nine reviews the findings of that research within the theoretical context of the literature. Overall, this thesis aims to address the following question: Is verbal behaviour functional in the formation and generalisation of equivalence classes and, if so, what are the parameters of that functionality?

1.2 BACKGROUND

A long history of comparative research has contributed to an assumption of radical behaviourism's inability to account for complex areas of human function--especially linguistic activity and untrained behaviour. A need for redress has contributed perhaps to the considerable interest that the study of stimulus equivalence has roused within the behavioural community over the last thirty years. From the outset, equivalence research has majoritatively employed human participants, the paradigm's raison d'être being the explanation of untrained higher-order behaviour. Its core theoretical debates have long centred on the nature and function of language, and its aims are regarded by many as a natural extension of decades of research into rule-governance. The questions remain, however: What is stimulus equivalence, and what is its relation to verbal behaviour?

1.3 THE STIMULUS EQUIVALENCE PARADIGM

1.3.1 Genesis

Latter-day enquiry into stimulus equivalence originated within a programme of research into reading comprehension among participants with intellectual disabilities (Sidman, 1971). The subject of this initial study was a severely intellectually-disabled adolescent boy, able to respond appropriately to spoken words and pictures, but not so to comprehend written words. When match-to-sample procedures (see Section 3.1) were used to establish a set of twenty conditional relations between spoken words and the pictures they represented, however, and another twenty between those spoken words and their written form, an equal number of bidirectional relations also emerged, untrained, between the written words and the pictures they represented (see Figure 1). This was an unprecedented finding, and one apparently unpredicted by simple conditioning principles (see Section 2.3). It was, however, a finding that was to prove highly replicable.

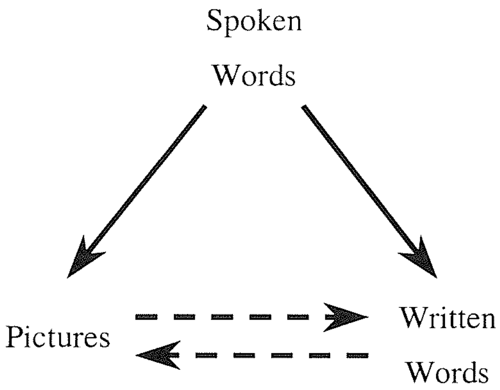

Figure 1. Sidman's (1971) equivalence paradigm. Arrows point from sample to comparison stimuli. Solid arrows represent trained relations, broken arrows represent emergent relations.

1.3.2 Development

The next eleven years witnessed a number of successful replications and extensions of these original findings, the majority of which continued to employ intellectually-disabled children and adolescents as participants (e.g., Constantine & Sidman, 1975; Dixon, 1977; Dixon & Spradlin, 1976; Gast, VanBiervliet, & Spradlin, 1979; Sidman & Cresson, 1973; Sidman, Cresson, & Willson-Morris, 1974; Spradlin, Cotter, & Baxley, 1973; Spradlin & Dixon, 1976). Conclusions from these studies, however, were phrased largely in the language of the then current, and now redundant, paired-associate research tradition (e.g., Jenkins, 1963; Jenkins & Palermo, 1964) which made heavy reliance on concepts of transfer of function and mediated generalisation via verbal response.

As Sidman (1994) has pointed out, such "response chaining" models fall victim to the limitations inherent in the experimental and theoretical practices of methodological behaviourism--not least through their failure consistently to predict experimental findings (i.e., Jenkins, 1965). Furthermore, such models tend to direct research attention away from the direct environmental stimulus-stimulus relations that have provided the stable basis of replicability characteristic of recent enquiry into derived relations (see Section 1.3.3). For similar reasons, and on grounds of parsimony, early formulations expressed in terms of participants' receptive and expressive speech have subsequently been discarded by most researchers (see Section 2.3).

1.3.3 A Mathematical Definition

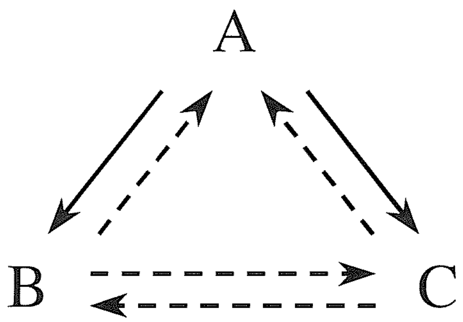

Figure 2. Sidman and Tailby's (1982) equivalence paradigm. Arrows point from sample to comparison stimuli. Solid arrows represent trained relations (AB and AC), broken arrows represent emergent relations of symmetry (BA and CA), transitivity (BC), and equivalence (CB).

As Sidman (1994) has pointed out, the terms "stimulus equivalence" and "equivalence relation" had been employed without precise definition in the research presented above--synonymically, in fact, for the phrase "substitutable for". It was not until 1982 that the mathematical definition of stimulus equivalence was proposed that has dominated subsequent research (Sidman & Tailby, 1982). In mathematics and logic, the concept of an equivalence relation requires that the three relational properties of "reflexivity", "symmetry", and "transitivity" hold between the elements of a given set. Suppose that a set consists of elements *a*, *b*, and *c*. If a relation **R** exists between elements *a* and *b* (*a***R***b*), and between elements *b* and *c*, then set theory requires that reflexive relations will exist between those elements (i.e., *a***R***a*, *b***R***b*, and *c***R***c*), that symmetric relations will exist also (i.e., *b***R***a* and *c***R***b*), and that a transitive relation will additionally hold between elements *a* and *c* (*a***R***c*). If this is so, then the inference will be valid that all of the elements within the set are equivalent--that they participate in an equivalence relation.

By its behavioural analogue (see Figure 2), in the case of the experimental paradigm outlined above (Sidman, 1971), reflexivity describes the conditional relation between a spoken word (A), picture (B), or written word (C) and itself (i.e., AA, BB, or CC relations, respectively). When conditional relations are established between A and B, and between A and C, symmetric relations are those that emerge, without further training, between C and A, and between B and A when samples and comparisons are reversed (i.e., BA and CA relations). For equivalence to be inferred, relations of transitivity and combined symmetry and transitivity (Sidman, Wynne, Maguire, & Barnes, 1989) must also form in the absence of training between B and C (i.e., BC and CB relations, respectively). The latter relations are often referred to simply as "equivalence" relations, in that their emergence requires the presence of both symmetric and transitive conditional relations (Sidman, 1990). As Sidman (1994) has observed, this formulation made possible research that focused on direct relations between stimuli (stimulus-stimulus relations), rather than those between stimuli and hypothetical mediating responses (stimulus-response relations). Stimulus equivalence had emerged from its origins in research into reading comprehension among participants with intellectual-disabilities to supply the theoretical potential for a general account of untrained human behaviour.

1.4 SUMMARY

Behaviour analysis has often attracted criticism for its supposed inability to predict and account for untrained behaviour. Findings of the orderly emergence of untrained behaviour from specific patterns of training have, however, led to the suggestion that such behaviour can be reliably predicted within an environmentally-based account of stimulus-stimulus relations. Although earlier theoretical attempts to explain emergent behaviour focused on stimulus chaining by mediational verbal responses, such accounts were rejected both because of their lack of parsimony and because of their failure to predict empirical findings. The instantiation of a mathematically based theory afforded the potential for a rigorous and coherent behavioural account expressed in terms of observable prior training and the testing outcomes thus generated. Much subsequent research has set out to explore the generality of the theory in terms of the participant populations and stimuli to which it can apply.

2. THEORETICAL REVIEW

This chapter aims firstly to document the generality of the stimulus equivalence paradigm by presenting a review of the stimuli and participant populations to which it can be applied (Section 2.1). Secondly, theoretical considerations pertaining to equivalence classes and their formation are presented in Section 2.2. Thirdly, the theoretical revisions of Sidman's (1994) formulation of equivalence are described briefly in Section 2.3, with regard to the literature previously reviewed. Lastly, this chapter aims to present a description of two major language-based accounts of emergent human behaviour that have been developed in recent years (Section 2.4).

2.1 GENERALITY OF THE PARADIGM

The generality of stimulus equivalence as an account of untrained behaviour has been extended by the publication of a large number of studies that have demonstrated equivalence phenomena using a wide variety of stimuli, presented across modalities, to a diversity of participants. The former aspect of the literature is reviewed in Section 2.1.1 and the latter is reviewed in Section 2.1.2.

2.1.1 Stimuli

Early research into equivalence employed mainly auditory and visual stimuli. Many studies (e.g., Constantine & Sidman, 1975; Sidman, 1971; Sidman & Cresson, 1973; Sidman et al., 1974) presented auditory and/or written stimulus names to their participants, in addition to pictures of the easily recognisable items to which those names referred. Similarly, Gast et al. (1979) employed auditory and written number names as well as their numeric counterparts. The first studies to employ purely arbitrary visual stimuli (i.e., abstract shapes), and purely arbitrary auditory stimuli (i.e., nonsense syllables) to control their class-membership (Dixon 1977; Dixon & Spradlin, 1976; Spradlin et al., 1973; Spradlin & Dixon, 1976) were also the first to extend the scope of equivalence beyond reading comprehension. These studies additionally set a methodological precedent that has informed the majority of subsequent research. Although the stimuli used were initially selected only to preclude pre-experimental stimulus associations among the intellectually-disabled participants to whom they were presented, such stimuli have additionally offered researchers the potential to control for adult participants' naming of experimental stimuli (see Section 2.4.2).

Commonly used abstract visual stimuli have included Greek or Cyrillic letters (e.g., Car, 1997; Devany, Hayes, & Nelson, 1986; Eikeseth & Smith, 1992; Holth & Arntzen, 1998; Lazar & Kotlarchyk, 1986; Sidman, Kirk, & Willson-Morris, 1985; Sidman & Tailby, 1982; Sidman, Willson-Morris, & Kirk, 1986) and abstract shapes and forms (e.g., Dube, McIlvane, Maguire, Mackay, & Stoddard, 1989; Fields, Adams, Verhave, & Newman, 1990; Fields, Reeve, Adams, & Verhave, 1991; Kennedy & Laitinen, 1988; Lynch & Green, 1991; MacDonald, Dixon, & LeBlanc, 1986; Saunders, Drake, & Spradlin, 1999; Steele & Hayes, 1991; Stromer & Osborne, 1982). Selection of such stimuli must be undertaken with care, however, as Stikeleather and Sidman (1990) have pointed out: Even abstract shapes can provide bases for consistent stimulus selection owing to unintended identity relations between stimuli (e.g., shared shape elements). Other extraneous similarities (e.g., shapes that look like parts of a car) should also be avoided for similar reasons.

Other research using visual stimuli has set out to capitalise upon for participants' verbal behaviour during experimentation, however, by the use of a variety of familiar, or easily nameable images (e.g., Bentall, Dickins, & Fox, 1993; Bentall, Jones, & Dickins, 1999; Cowley, Green, & Braunling-McMorrow, 1992; Dickins, Bentall, & Smith 1993; Holth & Arntzen, 1998; Mandell & Sheen, 1994; Smith, Dickins, & Bentall, 1996). Bentall et al. (1993), for example, presented three different types of visual stimuli to their adult participants, designated as preassociated pictograms (e.g., cactus, rose, palm), nameable nonassociated pictograms (e.g., cup, cross, book), and hard to name stimuli (abstract shapes), in order to evaluate potential differences in equivalence class formation as a function of naming. Mandell and Sheen (1994) likewise employed three classes of textual stimuli (i.e., phonologically correct pseudowords, phonologically incorrect pseudowords, and punctuation marks) to investigate whether the emergence of equivalence would be predictable as a function of the pronounceability of the stimuli used.

In the auditory modality, the most commonly used abstract stimuli have been nonsense words or syllables (e.g., Buffington, Fields, & Adams, 1997; Cullinan & Barnes, 1998; Dymond & Barnes, 1997; Fields et al., 1993; Saunders, Wachter, & Spradlin, 1988; Watt & Barnes, 1993), or tones (e.g., Bush, Sidman, & de Rose, 1989; Lane, Clow, Innis, & Critchfield, 1999). Although only one study has demonstrated equivalence using purely auditory stimuli (Dube, Green, & Serna, 1993), a large number of studies have done so using both auditory and visual stimuli (e.g., Bush, 1993; Green, 1990; Lipkens, Hayes, & Hayes, 1993; Lynch & Green, 1991; Saunders, Saunders, & Spradlin, 1990; Spradlin & Dixon, 1976). Overall, research has indicated that match-to-sample preparations employing auditory stimuli as samples and visual stimuli as comparisons are more likely to foster the emergence of equivalence than those employing purely visual stimuli throughout (Stromer & Mackay, 1996). It has been suggested, however (Dube et al., 1993), that the use of purely auditory stimuli may promote the emergence of equivalence among non-human participants (see Section 2.1.2).

Equivalence has been demonstrated to emerge between coin values and combinations (McDonagh, McIlvane, & Stoddard, 1984), between colours and colour names and numbers and number names (Mackay, 1985; Mackay & Sidman, 1984), between fractions and their decimal expression (Lynch & Cuvo, 1995), and between stimulus quantities (Kennedy & Serna, 1995). Equivalence has also been demonstrated between words in different languages (Joyce & Joyce, 1990; Sigurdardottir, 1992), between musical stimuli (Hayes, Thompson, & Hayes, 1989; Tena & Velazquez, 1997), using emotionally charged and emotionally neutral stimuli (e.g., Plaud, 1995; Plaud, Gaither, Franklin, Weller, & Barth, 1998; Plaud, Gaither, Weller, Bigwood, Barth, & von Duvillard, 1998), and among a variety of sexual stimuli (e.g., Barnes & Roche, 1997; Grey & Barnes, 1996; Roche & Barnes, 1996a, 1997; Roche, Barnes, & Smeets, 1997). A number of studies have also demonstrated the emergence of equivalence between elements of multi-element stimuli (see Section 2.2.6).

Regarding stimuli from other modalities, equivalence classes have been demonstrated to form incorporating gustatory stimuli (Hayes, Tilley, & Hayes, 1988), haptic stimuli (Tierney, De Largy, & Bracken, 1995), olfactory stimuli (Annett & Leslie, 1995), tactile stimuli (Belanich & Fields, 1999; Bush, 1993; O'Leary & Bush, 1996), between interocepetive and exteroceptive stimuli (DeGrandpre & Bickel, 1993; DeGrandpre, Bickel, & Higgins, 1992), and involving temporal differentiation (Rehfeldt & Hayes, 1998b). Although a sizeable literature confirms the cross-modal transfer of equivalence involving such stimuli (e.g., Augustson & Dougher, 1997; Lane et al., 1999; Sidman, 1971; Sidman & Cresson, 1973; Stromer & Mackay, 1996), little research has addressed quantitative issues affecting the formation of equivalence classes involving stimuli from different modalities (but see Belanich & Fields, 1999).

Harrison and Green (1990) have additionally demonstrated the formation of equivalence in the absence of reinforcing stimuli, by the presentation of constant sample-correct comparison pairings in the presence of inconstant incorrect comparison pairings. Saunders, Saunders, Kirby, and Spradlin (1988) have further reported both the development and merger of equivalence classes in the absence of reinforcing stimuli among participants with histories of equivalence class formation. Additional research has demonstrated the inclusion of reinforcing stimuli (Dube, McIlvane, Mackay, & Stoddard, 1987; Dube et al., 1989; Dube, Rocco, & McIlvane, 1989; McIlvane, Dube, Klederas, de Rose, & Stoddard, 1992) in equivalence classes, and the transfer of reinforcing and punishing stimulus functions via equivalence (e.g., Greenway, Dougher, & Wulfert, 1996; Hayes, Kohlenberg, & Hayes, 1991).

2.1.2 Participants

Another aspect of the equivalence paradigm's generality has been indicated by demonstrations of equivalence class formation across a wide range of participant populations. As mentioned above, early equivalence research employed intellectually-disabled participants (Constantine & Sidman, 1975; Dixon 1977; Dixon & Spradlin, 1976; Sidman, 1971; Sidman & Cresson, 1973; Sidman et al., 1974; Spradlin et al., 1973; Spradlin & Dixon, 1976), and many subsequent studies have likewise employed intellectually-disabled children, adolescents, or adults as participants (e.g., Bonta & Waters, 1981; Brady & Saunders, 1991; Dube et al., 1989; Glat, Gould, Stoddard, & Sidman, 1994; Haring, Breen, & Laitinen, 1989; Kennedy, Itkonen, & Lindquist, 1994; Lowenkron, 1988; Mackay, 1985; Mackay & Ratti, 1990; Saunders et al., 1990; Saunders & Spradlin, 1990; Stromer & Mackay, 1992a, 1992b; Stromer, Mackay, Howell, McVay, & Flusser, 1996; Zygmont, Lazar, Dube, & McIlvane, 1992). A number of other studies have compared intellectually-disabled and non intellectually-disabled participants of differing verbal abilities, further to investigate the role of verbal behaviour in equivalence class formation. As has been noted, however, such studies can, by their nature, provide only correlational data (Horne & Lowe, 1996).

Devany et al. (1986) compared the performance on equivalence tasks of mainstream verbally able children, verbally able children with intellectual-disabilities, and non verbally able children with intellectual-disabilities. Their findings indicated that although all children in the former two categories demonstrated the emergent relations of equivalence, the children in the latter category did not. Likewise, Eikeseth and Smith (1992) reported that equivalence classes did not form among the autistic children who participated in their study until verbal interventions taught them labels for the abstract stimuli with which they were presented. Even following this intervention, the least verbally able child tested did not demonstrate full equivalence. Barnes, McCullogh, and Keenan (1990), in an extension of the findings of Devany et al. (1986), compared the match-to-sample performance of mainstream children with that of verbally able and verbally disabled children with hearing deficits. Their findings indicated a far greater degree of consistency in equivalence class formation among the verbally able children, affording the conclusion that equivalence and verbal behaviour are closely related.

Although such studies have suggested overall correlations between verbal behaviour and equivalence class formation, they have nevertheless attracted criticism on the grounds that deficits in language development may well imply other physiological and experiential deficits which could equally explain equivalence failures without recourse to verbal behaviour (Sidman, 1994; Stromer & Mackay, 1996). Eikeseth and Smith's (1992) study has received additional criticism because of its potential confounding of verbal intervention with further exposure to the match-to-sample contingencies of the experiment (Mandell & Sheen, 1994; Stromer, Mackay, & Remington, 1996).

Although participants with learning disabilities have formed the basis of much equivalence research--both to explore the parameters of the paradigm's application and because of their relevance to the debate over equivalence and verbal behaviour--the majority of equivalence studies have nonetheless employed verbally able adults or children as participants. Such participants are most readily available for experimentation, and additionally demonstrate equivalence with the greatest facility. Findings from the numerous studies that have employed such participants have, however, fuelled debate regarding the necessity or sufficiency of verbal behaviour as a prerequisite for equivalence class formation. Another area of research to have stirred such controversy has been provided by the few empirical studies that have attempted to demonstrate equivalence using non-human participants.

The earliest comparative study to investigate the formation of bidirectional stimulus-stimulus relations (Sidman, Rauzin, Lazar, Cunningham, Tailby, & Carrigan, 1982) was also the first to suggest that equivalence might not emerge among non-humans. Following extensive training, the rhesus monkeys and baboons that participated in the experiment failed to demonstrate the emergence of even symmetric relations. This finding was further endorsed by Dugdale and Lowe (1990) who, using already highly verbally trained chimpanzees again found that, despite intensive visual-visual match-to-sample training, even symmetric relations failed to form. Failures to exhibit equivalence have also been reported using pigeons, monkeys, and chimpanzees as participants (e.g., D'Amato, Salmon, Loukas, & Tomie, 1985; Lipkens, Kop, & Matthijs, 1988; Rodewald, 1974; Yamamoto & Asano, 1995).

Although many researchers believe that an unequivocal demonstration of equivalence has not been achieved--and never will be--owing to non-humans' lack of verbal ability, a number of studies have nevertheless suggested that the constituent relations of equivalence can be demonstrated by pigeons (Kuno, Kitadate, & Iwamoto, 1994; Zentall & Urcuioli, 1993), monkeys (McIntire, Cleary, & Thompson, 1987), parakeets (Manabe, Kawashima, & Staddon, 1995), and sea-lions (Gisiner & Schusterman, 1992; Schusterman & Kastak, 1993, 1998; Schusterman, Kastak, & Reichmuth, 1997; Schusterman & Krieger, 1984). Such conclusions have not gone uncontested, however (e.g., Hayes, 1989; Horne & Lowe, 1996; Saunders, 1989), and it is generally accepted that an unequivocal demonstration of non-human equivalence has not yet been achieved. Nevertheless, Sidman (1994) has suggested that procedural, rather than species, characteristics may be responsible for non-human equivalence failures (see also Dube et al., 1993; Sidman, 1992; Zentall, 1998). Recent discussion of the facilitative effects of differential reinforcement and simple discrimination training (see Section 3.2.2) in the establishment of bidirectional relations among non-verbal participants has further supported this hypothesis (Saunders & Williams, 1998).

2.2 EQUIVALENCE CLASSES

The term "stimulus class" can be used to describe two or more stimuli that control the same class of response (Spradlin et al., 1973; Spradlin & Saunders, 1984). As Green and Saunders (1998) have observed, however, different types of stimulus class can be identified, whose definitions depend upon the formal properties of the stimuli of which they are composed. On this basis, a distinction can be drawn between "feature" classes and "arbitrary" classes (Stromer & Mackay, 1996).

Feature classes can be defined as classes composed of stimuli whose class membership is dependent upon physical similarity. Whether or not specific stimuli participate in such classes can be assessed as a function of primary generalisation gradients; i.e., on the basis of the probability with which a response, that has previously been occasioned by a given stimulus, will occur in the presence of novel stimuli whose membership of the same class is being assessed (Fields et al., 1991). Arbitrary classes, by contrast, may be composed of stimuli whose class membership does not depend on physical similarity, and which do not necessarily control the same behavioural function. Equivalence classes can be described as special instances of arbitrary classes, in that the relations between the stimuli of which they are composed must be reflexive, symmetric, and transitive. Although some studies have employed formally related stimuli to assay equivalence phenomena, the majority of studies assessing the formation and structure of equivalence classes have nevertheless employed formally unrelated stimuli.

A number of major aspects of research into equivalence class formation can be identified and are reviewed below. Studies which have investigated issues of class size are reviewed in Section 2.2.1, and studies assaying their expansion and merger are reviewed in Section 2.2.2. Investigations of the generalisation of equivalence classes are reported in Section 2.2.3, and a review of the literature relating to their formal structure is presented in Section 2.2.4. Studies which have investigated the contextual control of equivalence classes and equivalence between elements of multi-element stimuli are described in sections 2.2.5 and 2.2.6, respectively. Research into the delayed emergence and retention of stimulus equivalence is described in Section 2.2.7, and investigations of equivalence class formation on the basis of exclusion are reviewed in Section 2.2.8.

2.2.1 Size

As Fields and Verhave (1987) have observed, class size limits all other structural parameters of an equivalence class, and also determines the maximum number of derived relations within a class. Although early research into equivalence (Sidman, 1973; Sidman & Cresson, 1973; Sidman et al., 1974) demonstrated the emergence of classes composed of no more than three auditory or visual stimuli, subsequent studies have, however, set out to explore the parameters of equivalence with regard to the number of stimuli of which equivalence classes can be composed.

Sidman and Tailby (1982), in addition to their pioneering definition of equivalence, demonstrated the emergence of three four-member classes of equivalent abstract auditory and visual stimuli (i.e., Greek letters and letter names) using children as participants. Subsequently, the formation of five-member equivalence classes using arbitrary visual stimuli presented to adult participants has been demonstrated (Bentall et al., 1999), and six-member classes have also been shown to emerge using abstract visual stimuli (i.e., Greek letters) using adult participants (Sidman et al., 1985). Seven-member classes have been reported by Kennedy (1991), and Saunders et al. (1988) have further documented the establishment of control by abstract auditory stimuli over eight-member classes of abstract visual stimuli, using intellectually-disabled children and adults as participants.

Such findings have led some researchers (e.g., Sidman, 1994; Spradlin, Saunders, & Saunders, 1992) to propose that class-size may represent a variable that can affect both the formation and retention of equivalence classes (see Section 2.2.7). Other researchers (e.g., Adams, Fields, & Verhave, 1993a; Bentall et al., 1999; Fields, Adams, & Verhave, 1993; Fields et al., 1990; Fields, Reeve, Rosen, Varelas, Adams, Belanich, & Hobbie, 1997; Fields & Verhave, 1987; Fields, Verhave, & Fath, 1984), however, have addressed this issue with regard to the formal structure of equivalence relations, in terms of the nodal distance between stimuli (see Section 2.2.4). As Sidman (1994, p. 227) has observed, "if I were to hazard a prediction about the possible limits to the number of equivalent stimuli that a class could contain, I would guess that the existence of such a limit would become more likely as the number of nodes in the baseline conditional discriminations increased… but definitive experiments on class-size limitations have yet to be carried out."

2.2.2 Expansion and Merger

Green and Saunders (1998) have discussed a variety of ways in which the number of stimuli composing equivalence classes can be extended. Firstly, class "expansion" can be achieved by training a conditional relation between any member of an existing equivalence class and a novel stimulus, and extinction testing employed to assess whether the new stimulus has become equivalent to all other members of the pre-existing class. An early empirical demonstration of class expansion was provided by Dixon and Spradlin (1976), who established two classes of visual stimuli by employing every stimulus in each potential class in the role of both sample and correct comparison for every other prospective class member during reinforcement training. Subsequent training of responses to stimuli from those classes in the presence of novel auditory samples achieved class expansion. Similar demonstrations have been provided more recently using both mainstream and intellectually-disabled participants (e.g., Dube et al., 1987, 1989; Lazar, Davis-Lang, & Sanchez, 1984). Fields, Newman, Adams, and Verhave (1992) have further demonstrated class expansion using simple discrimination training (see Section 3.2.2). As Sidman (1994) has observed, however, the number of stimuli by which a class is expanded at one time, the size of the pre-existing classes to which those stimuli are added, and the number of nodal stimuli composing both the novel and pre-existing classes, may all quantitatively affect equivalence class expansion. Nevertheless, empirical confirmation of these suggestions has yet to be reported.

A related method by which the number of stimuli composing equivalence classes may be increased has been termed "class merger" (Sidman, 1994). For example, Sidman et al. (1985) established three 3-member equivalence classes composed of abstract visual stimuli, and another three equivalence classes composed of the same number of different abstract visual stimuli, using children as participants. Establishment of conditional relations between one set of stimuli from each class resulted in the emergence of six-member equivalence classes among five of the eight adults and children employed as participants. Similar findings have been reported by Saunders et al. (1988), although Sidman (1994) has subsequently noted that class merger may fail to occur as a function of contextual control (see

Section 2.2.5) by negative comparison stimuli, or as a result of conflict between class merger and the reinforcement contingencies of the experiment.

2.2.3 Generalisation

Not only can the number of stimuli composing equivalence classes can be increased through class expansion and class merger, but also through primary stimulus generalisation (Fields et al., 1997). As classes containing finite numbers of formally unrelated stimuli, equivalence relations have been described as arbitrary classes, in that they are not composed of the potentially unlimited numbers of physically similar stimuli definitive of feature classes (Stromer & Mackay, 1996). That relationships can exist between the two types of class has been suggested by demonstrations of the generalisation of equivalence classes to novel, yet physically similar, stimuli.

Haring et al. (1989), for instance, employed intellectually-disabled adolescents as participants to demonstrate the generalisation of equivalence classes to novel visual stimuli that shared either a high or moderate level of physical similarity with the visual stimuli used during baseline training. All their participants, however, demonstrated generalisation more readily when stimuli that shared a high level of physical similarity with the baseline stimuli were employed. Although a demonstration of the generalisation of identity matching to novel stimuli using pre-school children has been provided by Brown, Brown, and Poulson (1995), the majority of research into the formation of generalised equivalence classes has nonetheless been undertaken by Fields and colleagues (Adams, Fields, & Verhave, 1993b; Fields, Adams, Buffington, Yang, & Verhave, 1996; Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields et al., 1991; Meehan & Fields, 1995).

Fields et al. (1991), for example, established two three-member equivalence classes composed of auditory nonsense syllables and sets of visual stimuli composed of either "short" or "long" lines. When the equivalence classes had been established, novel lines of differing lengths were substituted as comparisons for the lines presented during baseline training and emergent testing. The results indicated that, in general, the likelihood of participants' choosing a given novel comparison was an inverse function of the difference in length between the novel lines and the lines used in training. Similar findings have also been also reported more recently utilising stimulus materials and participant populations similar to those described above (Fields et al., 1996, 1997; Meehan & Fields, 1995). Such findings have led to the suggestion that the generalisation of equivalence classes, and their merger with perceptual (i.e., feature) classes, may provide an empirical basis for predicting the emergence of both "natural kind" and "fuzzy superordinate" categories (Fields et al., 1996, 1991).

2.2.4 Structure

Fields and colleagues have not limited themselves to the study of equivalence class generalisation and the formation of natural categories, however. Variables affecting the speed and accuracy with which equivalence classes form have additionally been suggested by their exploration of the structure of equivalence classes in terms of "nodality", or the "nodal distance" between the stimuli of which those classes are composed (Adams et al., 1993a, 1993b; Fields et al., 1993, 1990, 1991, 1984; Fields, Adams, Verhave, & Newman, 1993; Fields & Verhave, 1987).

Whereas "singles" can be defined as stimuli that are linked through training to only one other stimulus, "nodal stimuli" or "nodes" can be defined as stimuli that are related by training to two or more other stimuli (Fields & Verhave, 1987). For example, if conditional discriminations are established sequentially between stimuli A, B, C, and D, reflexive (i.e., AA, BB, CC, and DD) and symmetric (i.e., BA, CB, and DC) relations are those between stimuli separated by zero nodes. The transitive relations that would also be expected to emerge during emergent testing would, however, be composed of stimuli separated by one- and two-nodes' distance (i.e., AC and AD relations, respectively). Likewise, the nodal distance between stimuli related by combined symmetry and transitivity (or equivalence) would be respectively the same (i.e., one-node CA and two-node DA relations). Sequential training involving incrementally greater numbers of stimuli would necessarily require that those stimuli be separated by incrementally greater nodal distances.

Fields and Verhave (1987) have additionally suggested that the nodal distance separating stimuli will determine the control exerted by one stimulus over another during emergent testing. Fields et al. (1990) confirmed this hypothesis by providing an empirical demonstration that class-consistent comparison selection among adult participants decreased as a function of the nodal distance between the visual nonsense syllables employed as stimuli. Those researchers also observed, however, that class-consistent responding on trials involving greater nodal distances between stimuli increased with repeated testing (see Section 2.2.7). Dube et al. (1993) have subsequently reported similar findings using four-member classes composed of auditory stimuli, and Kennedy (1991) has likewise documented effects of nodal distance within seven-member equivalence classes composed of arbitrary visual stimuli using adult participants.

Effects of nodal distance have not only been evident in the accuracy with which participants perform on equivalence tasks, however. Wulfert and Hayes (1988) were the first researchers to assess whether effects of nodal distance could be observed through measurement of participants' response latencies during emergent testing (see Section 3.3.5). When three-member equivalence classes composed of arbitrary visual forms had been established among their adult participants, response latencies were observed to be greater on trials of one-node transitivity than on trials of baseline relations. That participants' response latencies increase as a direct function of the nodal distance separating stimuli has been confirmed by a number of other studies. Bentall et al. (1993), for instance, employed a variety of arbitrary and non-arbitrary visual stimuli to demonstrate that their adult participants' response latencies were significantly greater on trials of one-node transitivity and equivalence than on trials of baseline relations and symmetry. These researchers termed this finding the "transitivity latency effect". A greater number of non class-consistent responses were also observed on the former trial types. This finding was termed the "transitivity error effect". As previously suggested by the findings of Fields at al. (1990), however, both effects were seen to diminish with repeated testing in extinction. Such findings have subsequently been confirmed by other researchers (Bentall et al., 1999; Dickins et al., 1993; Spencer & Chase, 1996). Investigations of the potential interactions between training and testing protocols and the effects of nodal distance are reviewed below (see sections 3.1.1 and 3.1.2, respectively).

Although research has so far supported the utility of nodal distance as an explanatory concept, Sidman (1994, p. 539) has objected to the overtones of structuralism inherent in the term's use. To avoid these, and implications of a "linear-associative conception of stimulus control", he has suggested substitution of the term "nodal number". Most researchers have, however, retained use of Fields and colleagues' original terminology.

2.2.5 Contextual Control

The term "stimulus control" describes the control of behaviour by its environment. Although operant conditioning is traditionally explained in terms of the three-term contingency of antecedents, behaviour, and consequences, such behavioural relations can, themselves, come under conditional control. The four-term (or first-order) behavioural relations thus described can, however, themselves be brought under conditional control, providing a fifth term to the contingency that is usually described as second-order, or contextual, control. Although theoretical discussions of relationships between equivalence and contextual control were provided by Sidman et al. (1985) and Sidman (1986), the first empirical study to investigate the emergence of contextually controlled equivalence classes was reported by Lazar and Kotlarchyk (1986). Using children as participants, these researchers demonstrated that equivalence relations composed of sequence-classes (see Section 3.2.5) of abstract visual stimuli could be brought under the contextual control of tones.

A large number of empirical studies have subsequently demonstrated the contextual control of equivalence relations and, in so doing, have employed a wide variety of participants and experimental stimuli. Kennedy and Laitinen (1988), for instance, used purely abstract visual stimuli presented to adult participants to assess the effects of training order (see Section 3.1.1) on the formation of contextually controlled equivalence classes. Although these researchers concluded that the non-emergence of transitive relations observed when five-term contingencies were established prior to four-term contingencies was either because of overshadowing (Honig & Urcuioli, 1981; Mackintosh, 1977) or because of participants' use of verbal rules, procedural limitations rendered the findings of this experiment inconclusive (see Section 2.2.8). Another study, again employing adult participants, but using arbitrary visual stimuli under the contextual control of tones (Bush et al., 1989), was the first to document the delayed emergence (see Section 2.2.7) of contextually controlled equivalence classes, and other studies have since confirmed this finding (e.g., Hayes et al., 1991; Meehan & Fields, 1995).

Gatch and Osborne (1989), having noted that single stimuli had been employed as contextual cues in all previous research, provided a demonstration that classes of stimuli, formed at the level of the five-term contingency, can control responding at the four-term level. Two classes of equivalent abstract visual stimuli were established using adult participants, with class membership determined by the presence of one or other of two additional visual stimuli (i.e., contextual stimuli). When these classes had been established, novel visual stimuli were related to each of the two previously established contextual stimuli, forming classes of contextual stimuli that subsequently controlled participants' first-order conditional discriminations. Extending these findings, Lynch and Green (1991) established contextual control by tones over classes of visual stimuli, and then trained novel visual stimuli to the tones, thus providing a demonstration of the cross-modal transfer of contextually controlled equivalence classes among adult participants.

Other studies have investigated the formation of contextually controlled equivalence classes with regard to the transfer of functions (see Section 2.4.1). Hayes et al. (1991), for example, demonstrated that consequential functions can transfer through equivalence relations on the basis of contextual cues, using arbitrary visual stimuli presented to adult participants. Kohlenberg, Hayes, and Hayes (1991), likewise using arbitrary visual stimuli but also visual proper names, demonstrated that contextual control can itself transfer through equivalence classes among adult participants, thus suggesting a possible model of social stereotyping. Barnes, Browne, Smeets, and Roche (1995) have subsequently extended these findings by providing a demonstration of the transfer of functions using auditory and visual stimuli presented to three- to six-year-old children.

Employing adolescent participants, Steele and Hayes (1991) pre-trained arbitrary visual stimuli as contextual cues to control relational responding to additional arbitrary visual stimuli. The performances observed led to the conclusion that non-arbitrary stimulus relations (i.e., "same", "opposite", and "different") can be brought under contextual control and additionally applied to novel sets of both formally related, and formally unrelated, stimuli. Roche and Barnes (1996a) have further extended this paradigm to provide a possible model of sexual categorisation, using visual stimuli presented to adult participants. Meehan and Fields (1995) have also provided a demonstration of the generalisation of contextually controlled equivalence classes of arbitrary visual stimuli to other such stimuli as a function of physical similarity. The emergence of contextually controlled equivalence classes as a result of exclusion (see Section 2.2.8) has also been demonstrated using arbitrary visual stimuli presented to adult participants (Meehan, 1995).

Although further research (Wulfert et al., 1994) has demonstrated the formation even of third-order equivalence classes (i.e., second-order equivalence classes themselves under contextual control) using arbitrary visual stimuli presented to adult participants, only one study (Carr, 1997) has so far attempted to assay the relationships between the formation of contextually controlled equivalence classes and participants' verbal behaviour during experimentation (see Section 5.1). Furthermore, although many equivalence studies have employed both visual stimuli and auditory tones as contextual cues, no evidence is yet available as to whether any particular type of contextual cue is more or less efficacious in promoting the emergence of contextually controlled equivalence classes.

2.2.6 Stimulus Compounds

As various researchers have pointed out (Maguire, Stromer, Mackay, & Demis, 1994; Schenk, 1993), the majority of equivalence studies have only employed "simple" (i.e., uni-element) stimuli as samples and comparisons. Nevertheless, such stimuli are seldom encountered by humans outside the laboratory, and the demonstration of emergent relations between elements of "stimulus compounds" (i.e., complex or multi-element stimuli) would therefore serve to endorse the ecological validity of equivalence as an account of untrained behaviour (Stromer & Stromer, 1990a). In many laboratory preparations, multi-element sample stimuli have been composed of the arbitrary simple stimuli used as comparisons (Stromer & Mackay, 1990). For example, on successive two-choice discriminations in which the correct comparison is either a colour or form, and the incorrect comparison an extraneous colour or form of the same nature as the correct comparison, the sample would be composed of a colour and form superimposed. Additional procedures, however, have also been employed.

As Stromer, McIlvane, and Serna (1993) have noted, studies that have engendered observation of elements of complex stimuli among their participants have provided repeated evidence of the formation of equivalence relations among those elements. Stromer and Stromer (1990a, 1992), for example, presented multi-element sample stimuli, composed of combinations of differing colours and unvarying tones, and arbitrary visual forms as comparisons. Comparison selection was found to be dependent upon the combination of the auditory and visual stimulus elements presented as samples. The results indicated not only that the individual sample elements had become correlated with comparison selection but also that equivalence classes had been established, because the tones presented were common across trials. An additional finding reported was that although the contingencies had not required participants to attend to both sample elements on every trial, relations had nevertheless been established between comparisons and the ostensibly redundant elements of the sample stimuli. These findings have been supported by those of Stromer and Stromer (1990b).

Other researchers have extended these findings in a number of ways. Stromer and Mackay (1992), for example, used delayed identity matching procedures involving complex samples composed of easily recognisable pictures and those pictures' names presented visually to establish rudimentary spelling skills among intellectually-disabled participants. Stromer and Mackay (1993) have replicated this finding using intellectually-disabled adolescents as participants, as have Maguire et al. (1994), who employed both adults with autism and mainstream children. Maguire, Stromer, and Mackay (1995) have further indicated that delayed match-to-sample procedures (see Section 3.1) may facilitate the emergence of equivalence between elements of complex stimuli, owing to such procedures' reduction of the potential for control by contiguously presented, yet redundant, sample stimulus elements.

Further to extend the scope of enquiry, Schenk (1993) employed identity matching procedures to establish equivalence among elements of complex visual stimuli (i.e., colours and abstract forms) and simple comparisons, using children as participants. Smeets, Schenk, and Barnes (1994) have subsequently endorsed Schenk's (1993) conclusion that the formation of equivalence observed had resulted from the requirement to respond differentially to the colour-form elements of the multi-element stimuli presented as samples. A demonstration of errorless transfer from identity to arbitrary match-to-sample has also been provided by Smeets and Striefel (1994), using arbitrary visual multi-element stimuli presented to children. The emergence of equivalence involving elements of complex stimuli has additionally been demonstrated by Schenk (1995), using no-reinforcement simple discrimination tasks. More recently, Rehfeldt, Dixon, Hayes, and Steele (1998) have provided evidence of the blocking effect (e.g., Kamin, 1969) in the formation of equivalence relations between elements of complex stimuli, although the variable nature of their findings did not afford definitive conclusions.

2.2.7 Delayed Emergence and Retention

Delayed emergence can be defined as the emergence of stimulus equivalence, subsequent to its non-emergence, as a result of repeated testing in extinction (Sidman, 1994). Although many studies have indicated that participants can demonstrate the emergent performances definitive of equivalence immediately subsequent to baseline training, Spradlin et al. (1973) were the first researchers to observe that, among their intellectually-disabled participants, equivalence emerged only after repeated extinction testing. Although such findings led initially to the use of the term "gradual emergence" as a description for such emergent behaviour, demonstrations that equivalence can emerge suddenly as a result of repeated testing in extinction (e.g., Sidman et al., 1985) have subsequently led most researchers to relinquish use of the term (Sidman, 1994).

A substantial literature documents the delayed emergence of equivalence among a variety of participant populations and using a variety of stimuli (e.g., Bush et al., 1989; Cowley et al., 1992; Harrison & Green, 1990; Lazar et al., 1984; Lazar & Kotlarchyk, 1986; Saunders et al., 1988; Sidman et al., 1986; Sigurdardottir, Green, & Saunders, 1990; Stromer & Osbourne, 1982), and has thus urged consideration of the phenomenon in any empirical investigation of equivalence (Stikeleather & Sidman, 1990). Holth and Arntzen (1998) have recently investigated the relationships between delayed emergence and the non-emergence of equivalence with regard to stimulus familiarity.

As Mackay (1991) has noted, observations of the "retention" (i.e., the long-term stability) of equivalence relations have suggested links with the study of human memory. The evidence upon which his statement was based, however, remains largely anecdotal and, to date, only two published studies (Saunders et al., 1988, 1990) have set out specifically to investigate the retention of equivalence relations. Only three others (Green, Mackay, McIlvane, Saunders, & Soraci, 1990; Mackay, 1991; Saunders et al., 1988) have discussed issues raised by that research, although one additional study (Mackay & Ratti, 1990) has suggested links between equivalence class formation and human short-term memory. Saunders et al. (1988) reported that auditory control of eight-member equivalence classes of arbitrary visual stimuli had "remained intact" for periods of two to five months among a number of the intellectually-disabled adults and children whom they had employed as participants. These researchers additionally noted the delayed re-emergence of equivalence classes among a number of their participants, and questioned whether class size and the number of classes previously established should not be regarded as critical variables in the long-term stability of equivalence relations. Spradlin, Saunders, and Spradlin (1992, pp. 33-34) have further suggested that class size may help to determine whether conditional relations between stimuli that have been forgotten will ultimately be remembered, stating that "as the number of stimuli within each equivalence class increases, the number of possible ways of recovering [those] relations is increased dramatically".

Empirical evidence of the retention of equivalence classes across two and three year re-test periods has been provided by Saunders et al. (1990), who employed an adult with mild intellectual-disabilities as a participant, and abstract visual forms as stimuli. Such long term stability suggested to these researchers strong links between equivalence class formation and verbal behaviour. Saunders et al. (1988) have further suggested that the differential deterioration of the constituent relations of equivalence over time may indicate qualitative differences between those relations (cf. Fields et al., 1990). These researchers have additionally noted that it "will be equally informative to learn whether the long-term retention of relations is also affected by whether or not the relations were developed in forced, two-choice procedures" (Saunders et al., p. 161), and it is to a discussion of research involving such preparations that we now turn.

2.2.8 Exclusion

A number of early equivalence studies employed match-to-sample procedures in which a sample was presented with only two comparison stimuli (e.g., Dixon & Spradlin, 1976; Sidman, 1971; Sidman & Cresson, 1973; Spradlin et al., 1973), and other more recent research has done likewise (e.g., Kennedy & Laitinen, 1988; Mandell & Sheen, 1994; Saunders et al., 1988; Spradlin & Saunders, 1986). Use of such procedures has, however, highlighted both methodological and theoretical considerations in the conduct equivalence research, as Stromer (1986) has pointed out. Such considerations are usually addressed under the rubric of exclusion--the control of comparison selection by negative stimulus relations. Exclusion is also sometimes referred to as "S-" as opposed to "S+" control (e.g., Dixon, 1977), or as a "reject", rather than a "select" relation (Green & Saunders, 1998).

Conditional discriminations established by match-to-sample training can be described as "*if... then...*" relations, as Sidman (1994) has observed. For example, the contingencies operational within a two-choice procedure may foster learning that "*if* stimulus A1 is presented as sample, *then* B1 is the correct comparison, rather than B2". In the same example, exclusion can be described as learning that "*if* A1 is presented as sample, *then* the correct comparison will not be B2": As Saunders and Green (1998) have pointed out, two-choice match-to-sample procedures may often establish the latter type of stimulus control. Because the aim of equivalence research is usually to establish control by correct (i.e., class-consistent) comparisons, researchers may remain unaware that control by a negative comparison has been engendered by their experimental procedures, and confounds thus introduced into their research (Sidman, 1994). As Stikeleather and Sidman (1990) have pointed out, control by negative stimuli can imply simple error and, generating low accuracy responding, give the impression of failure on tests of equivalence. Equivalence may nonetheless have been established, however, albeit between stimuli other than those expected. These hypotheses have received empirical validation using mainstream adults and intellectually-disabled adolescents as participants (e.g., Dixon, 1977; Johnson & Sidman, 1993; Meehan, 1995; Stromer, 1986; Stromer & Osborne, 1982). Performance commensurate with the phenomena of exclusion has also been observed in the behaviour of a "symmetry-emergent" chimpanzee (Tomonaga, 1993).

2.3 THEORETICAL CONSIDERATIONS

The last twenty years have witnessed an explosive growth in the number of studies that have addressed equivalence phenomena. Although the overwhelming majority of these studies have reported findings commensurate with the predictions of stimulus equivalence, a number have nevertheless offered refinement to various aspects of the theory, or suggested its potential shortcomings as an account of untrained behaviour. As Horne and Lowe (1996, p. 228) have observed, "like many theoretical accounts, Sidman's has evolved to accommodate empirical findings that do not fit easily within his initial formulations". Sidman's (1994) responses to such considerations have further been described by these researchers as an "extraordinarily ambitious revision of existing behavioural theory." (Horne & Lowe, 1996, p. 228).

As noted above, research into stimulus equivalence evolved from attempts to provide a functional analysis of reading comprehension among intellectually-disabled children and adolescents (Sidman, 1971; Sidman & Cresson, 1973). By 1982, however, research had suggested a generality to the paradigm that warranted, and indeed necessitated, a thoroughgoing definition of equivalence. Whereas equivalence research had, until then, been presented largely in terms of constructs borrowed from paired-associate research into "stimulus equivalence" (Jenkins, 1963; Jenkins & Palermo, 1964), the findings of Sidman et al. (1974) were the last from that laboratory to be phrased in terms of "mediated transfer" and "stimulus equivalences", as Sidman (1994) has pointed out. Sidman and Tailby's (1982) seminal paper was the first to report emergent behaviour within the mathematical framework that has dominated subsequent research.

Findings that humans can demonstrate equivalence relations composed of six- and even eight-member classes of stimuli (Saunders et al., 1988; Sidman et al., 1985) suggested additional aspects of the equivalence paradigm's generality that led Sidman (1990) to propose that equivalence may denote a basic stimulus function--a behavioural "primitive", similar in conceptualisation to reinforcement and discrimination in traditional behavioural theory. In accordance with the basic tenets of behaviour analysis, however, Sidman (1994) has maintained that the formation of equivalence classes should nevertheless be regarded as the result of the prevailing contingencies of reinforcement.

A shortcoming of Sidman and Tailby's (1982) initial formulation of equivalence had been suggested by its failure to incorporate reinforcing stimuli within the account (Sidman, 1994), and data showing that relations including such stimuli can indeed form (Dube et al., 1987, 1989; McIlvane et al., 1992) led to the conclusion that any kind of objects or events can become related by equivalence. As Sidman (1994, p. 384) has stated, "an equivalence relation is made up of pairs of events, with no restriction on the nature of the events that make up the pairs. The locus of those events, whether it be in the... organism or [its] living or non-living environment, is irrelevant." Evidence that equivalence can emerge at the level of the three-term (e.g., Barnes & Keenan, 1993; McIlvane, Dube, Klederas, Iennaco, de Rose, & Stoddard, 1990; Sidman et al., 1990) and even the two-term contingency (e.g., Saunders & Spradlin, 1989; Schenk, 1995; Smeets & Barnes, 1997; Smeets, Barnes, & Luciano, 1996) led Sidman (1994, p. 386) to "consider a more intimate connection than [he] had heretofore suspected between reinforcement and equivalence", and further to suggest that equivalence may underlie both operant and respondent conditioned reinforcement. As Sidman (1994, p. 391) has noted, "one of the consequences of including equivalence relations among the basic outcomes of reinforcement... is that conditioned reinforcement then becomes derivable". Indeed, "we can derive both operant and Pavlovian conditioned reinforcement in the same way--as the result of the formation of equivalence relations and the merger of equivalence and reinforcer classes" (Sidman, 1994, p. 399). Extending his reasoning further, Sidman (1994, pp. 403-404) has suggested that "if the CS-UCS pairing creates an equivalence relation, then the establishment of the equivalence relation, rather than the creation of a new stimulus-response relation, can perhaps be taken as the defining feature of Pavlovian conditioning".

The inclusion of defined responses within equivalence relations brought with it the need for further theoretical revision, however, because Sidman (1994, p. 377) had "come up against the impossibility of completely separating... responses from stimuli". With regard to this consideration, Sidman (1994, p. 386) has asserted that "equivalence relations have their own defining characteristics, none requiring the stimulus/response dichotomy". According to Sidman (1994, p. 379), therefore, "the inclusion of responses not only permits but forces us to maintain the our set-theory definition". Further to increase the scope of his mathematical analogy, Sidman (1994) has additionally incorporated concepts of set union, intersection, and partition within his account: As Sidman (1997a, p. 141) has pointed out, "it is rare, probably even impossible, for any element to belong to just one class. When two or more classes have members in common, they may merge (set union) or remain independent (set intersection). Contextual factors will determine whether set union or intersection takes place". Nevertheless, "mathematics does not predict when or how control by context arises.... Contextual control arises because the reinforcement contingencies permit or demand it" (Sidman, 1994, pp. 529-530).

Regarding the assertion that equivalence may underlie conditioned reinforcement at the level of both the two- and three-term contingency, Sidman (1994, p. 408) has indicated that his "solution is... drastic. One large equivalence class *must* emerge when the establishing contingencies share the same reinforcer and defined response.... This very reason for expecting failures imposes on us an obligation to explain not the occasional negative results but rather, to explain why the standard experiment usually succeeds in demonstrating equivalence [emphasis in original]". As Horne and Lowe (1996) have observed, however, this revision begs the question not of how equivalence relations form, but rather how "overarching" equivalence classes are broken down into the discrete relations reported in numerous equivalence studies. Sidman (1994) has suggested that such classes are broken down by a process of "selective dropping out"--although the nature of this process remains to be explained (Horne & Lowe, 1996).

Sidman (1994, p. 282) has addressed a further question, however: Is participants' "naming [of stimuli] more than facilitative? Is it *necessary*..." for the emergence of equivalence? Sidman (1994, p. 281) has summarised his views thus: "In questioning the role of naming in the formation of equivalence relations, we were concerned about the notion, analogous to the mediation hypothesis in the earlier paired-associate work, that subjects had to give all members of a class of equivalent stimuli the same name.... Human subjects are, of course, likely to name the stimuli that we present to them and naming may indeed foster original learning and even remembering... and may facilitate equivalence relations.... The reasons naming may serve these functions are not clear and deserve more research" (Sidman, 1994, p. 281). In recent years, other researchers have attempted to supply both empirical and theoretical resolution to this and other questions regarding the potentially functional role of verbal behaviour in stimulus equivalence.

2.4 ALTERNATIVE ACCOUNTS OF EMERGENT RELATIONS

Over the last twenty years, equivalence research has attracted the attention of an increasing number of behaviour analysts. Two researchers in particular, S. C. Hayes and C. F. Lowe, both with backgrounds of research into rule-governance, have formulated alternative accounts of emergent behaviour that have both focused on human participants' verbal behaviour, and its potentially determining role in the generation of stimulus equivalence. Within the last decade, both of these accounts have developed to form the basis of empirical and theoretical analyses of emergent behaviour and, as a result, debate has intensified with regard to the primacy of verbal behaviour or equivalence.

2.4.1 Arbitrarily Applicable Relational Responding

Hayes and colleagues have, in a number of papers (e.g., Hayes, 1991, 1994; Hayes & Hayes, 1989, 1992; Steele & Hayes, 1991), proposed that equivalence should be regarded not as a behavioural primitive, as Sidman (1990, 1994) has suggested, but rather as one of a number of derived stimulus functions which, along with other higher-order behaviour, result from prolonged exposure to the reinforcement contingencies operating within a verbal community.

Arguing that many species exhibit the ability to respond to non-arbitrary relations between objects and events (for example, "longer than" or "shorter than"), Hayes (1989) has suggested that humans' experiential history may include training of a kind that permits other types of "relational" responding to emerge. In learning to name an object, for example, a child may be taught to respond differentially to certain sounds (and later to produce them) in the presence of specific objects, and vice versa. Hayes (1989) has suggested that because this kind of bidirectional responding occurs only within certain contexts (in the example above, the context of naming), sufficient instances of directly trained symmetrical responding will cause the behaviours involved to emerge with respect to novel stimuli, given the relevant contextual cues. Such relations, it should be noted, are held to be essentially arbitrary in nature, being based not on any formal properties of the stimuli involved, but on the nature of the context in which they occur. These behaviours have therefore been referred to as "arbitrarily applicable relational responding", an epithet defined in terms of its three component properties of "mutual entailment", "combinatorial mutual entailment", and "transfer of functions" (Hayes, 1989, 1991).

Hayes (1989) has proposed that mutual entailment operates when the context brings to bear on a situation an individual's history of a particular type of relational responding. For example, if a particular type of relation is specified between two events (A and B) then another relation is necessarily entailed between B and A: If, for instance, A is smaller than B, then B must be larger than A. Symmetry may therefore be viewed as a special case of mutual entailment (Barnes, 1994), although as Sidman (1994) has observed, unlike symmetry, the AB and BA relations are not required to be functionally identical. Combinatorial mutual entailment describes the following relation: If A is related to B, and B is related to

C in a given context, then (with some similarity to the relations of transitivity and combined transitivity and equivalence in stimulus equivalence) relations are also entailed between A and C, and therefore between C and A. Once again, however, the relations described by combinatorial mutual entailment are not required to be functionally identical.

The final characteristic, transfer of functions, is that which "provides the psychological importance" of the events described above (Hayes & Hayes, 1992). If an event A has a psychological function, and there is a derived relation between event A and event B, then under certain conditions B may acquire a new psychological function, given that suitable contextual cues are present. This will be based on the function of A, and the relation between A and B. Mutual entailment and combinatorial mutual entailment are therefore themselves interpretable as transfers of functions in a limited sense, and arbitrarily applicable relations as patterns of the mutual transformation of both relational and non-relational stimulus functions (Hayes & Hayes, 1992).

A further term, "relational frame" was coined to designate particular kinds of arbitrarily applicable relational responding (Hayes & Hayes, 1989). A relational frame describes a type of responding that shows the three definitional properties of mutual entailment, combinatorial mutual entailment, and transfer of functions, and is therefore a term applied to particular patterns of the mutual transformation of stimulus functions (Lipkens et al., 1993). Hayes and Hayes (1989) list a variety of such frames, including those of "co-ordination", "opposition", "distinction", and "comparison". In this view, stimulus equivalence is a special case of the frame of "co-ordination" (Hayes & Hayes, 1992). A number of empirical studies have supported these hypotheses (e.g., Barnes & Keenan, 1993; Dymond & Barnes, 1994, 1995, 1996; Lipkens et al., 1993; Roche & Barnes, 1996; Steele & Hayes, 1991) and the account has been extended to provide a possible model of social stereotyping (Kohlenberg et al., 1991), and to account for relations involving gustatory (Hayes et al., 1989) and auditory musical stimuli (Hayes et al., 1989).

Another account of emergent behaviour, highlighting the primacy of language over stimulus equivalence, has been proposed by Dugdale and Lowe (1990), and most recently and comprehensively by Horne and Lowe (1996, 1997).

2.4.2 Naming and Other Symbolic Behaviour

During the late 1970s and 1980s, investigation was made into differences between the schedule-controlled behaviour of human adults, children, and non-humans (e.g., Bentall, Lowe, & Beasty, 1985; Lowe, 1979), and the conclusion reached that the development of verbal behaviour was responsible for the inter-species differences observed. Extending this approach to the field of stimulus equivalence, Dugdale and Lowe (1990) proposed that verbal behaviour sets the occasion for the emergence of equivalence, via participants' naming of the stimuli involved.

Contrary to Sidman's conceptualisation of equivalence as a behavioural primitive underpinning linguistic function (Sidman, 1990, 1994), Lowe and colleagues (Dugdale & Lowe, 1990; Horne & Lowe, 1996, 1997) have argued that equivalence, and other higher-order human behaviour, result from participants' overt or covert "naming" of stimuli, regardless of modality. In this view, naming is bidirectional "stimulus-classifying behavior" (Horne & Lowe, 1996, p. 227) and describes the fusion of speaker and listener behaviour resulting from an individual's history of reinforcement within a verbal community. The "name relation" is indicated as the basic unit of verbal behaviour, and success on tests of equivalence is proposed to result either from participants' common naming of individual stimuli or through their linking of individual stimulus names by intraverbal rules (Horne & Lowe, 1996).

Dugdale and Lowe (1990) were among the first researchers to propose that participants' verbal behaviour enables the emergence of equivalence. To use an example from that paper; although a fifty pence piece and five ten pence pieces offer very different perceptual cues, they may be accepted as equivalent if the common label "fifty" is applied to both. Additionally, even if the stimuli do not receive such a common name, it is proposed that equivalence may still result if the individual stimulus names are incorporated into a verbal rule (for example, "A means B", or "B goes with A"). As has been pointed out, however, this formulation leaves the "problem of accounting for how naming comes about and how it gives rise to... emergent or derived stimulus relations" (Horne & Lowe, 1996). To counter, a detailed and plausible developmental account of naming was proposed.

Drawing heavily on Skinner's (1957) account of verbal behaviour and a variety of findings from other branches of the discipline, Horne and Lowe (1996) have described the development of speaker and listener behaviours in the mainstream child from their genesis in echoic behaviour, through tacting and manding, and hence to complex intraverbal functioning. In this view, the name relation incorporates elements of all these classifications, enabling the development of receptive and expressive behaviour which, being shared with and developed by the verbal community, allows the child to interact with environmental objects and events in a socially appropriate way. As the fundamental unit of verbal behaviour, naming is described as a "circular" relation, exhibiting the bidirectional properties of speaker and listener behaviour which, resulting in symmetry between an object or event and its name, enable success on tests of equivalence.

Naming, whether common or intraverbally linked, is therefore presented as a necessary and sufficient precondition for the emergence of equivalence, regardless of the stimuli and modalities involved. Naming, however, "should not be viewed as *mediating* the establishment of stimulus classes: Naming *is* stimulus classifying behavior [emphases in original]" (Horne & Lowe, 1996, p. 226-227). It is allowed, however, that naming may both facilitate, or hinder the formation of equivalence relations, depending on whether the "intraverbal sequences that are formed before... testing are congruent with the experimenter defined classes" controlling the experiment (Horne & Lowe, 1996, p. 226). A number of empirical studies have lent support to this view, with findings reported that the teaching of stimulus names can both increase or reduce participants' ability to respond appropriately (e.g., Bentall et al., 1993; Dickins et al., 1993; Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Lowe & Beasty, 1987; Mandell & Sheen, 1994).

Mandell and Sheen (1994), for example, employed three classes of textual stimuli to evaluate the effects of naming among their undergraduate participants; phonologically correct, pronounceable pseudowords (e.g., SNAMB), phonologically incorrect pseudowords (e.g., NSJBM), and punctuation marks (e.g., +]*^!). If, as they suggested, naming is an important determinant of class formation, then the pronounceability of presented stimuli should prove indicative of the speed and accuracy with which classes form. Experiment 1 confirmed that participants exposed to pronounceable stimuli demonstrated equivalence more quickly and with greater consistency than participants in the other conditions, and also showed that participants in those conditions tended spontaneously to produce idiosyncratic names for the unpronounceable stimuli. A second experiment indicated that when participants were pre-trained to apply names orally to phonologically incorrect pseudowords, their performance was enhanced in comparison to participants who received no such pre-training.

Participants' verbal behaviour should not be regarded as a panacea for remediating equivalence failures, however, as Horne and Lowe (1996) have observed. In addition to evidence that some verbally able participants fail equivalence tests (e.g., Dugdale & Lowe, 1990; Eikeseth & Smith, 1992), other research has indicated that participants' verbal behaviour can either facilitate or hinder class formation, depending on the congruence of the naming strategies employed with the experimenter-designated classes governing positive test outcomes. Dickins et al. (1993) first taught three groups of participants baseline relations between sets of pictograms, after which participants in two of the groups were taught paired associations between the names they had given those stimuli forming verbal associations discordant with the stimulus classes already established by match-to-sample training. Participants in the third group were taught paired associations between neutral names and those of the experimental stimuli. Participants exposed to across-class paired associations were less successful on subsequent tests of equivalence than those in the latter group, with baseline discriminations repeatedly superseded by their subsequent verbal training.

Bentall et al. (1993, Experiment 1) presented three groups of undergraduate participants with different types of visual stimuli designated as preassociated pictograms, nameable nonassociated pictograms, and hard to name abstract stimuli. As predicted, equivalence was demonstrated most quickly and with fewest errors by participants exposed to the preassociated stimuli and most slowly by those in the abstract condition, although post-experimental interviews failed to reveal the consistent use of class names. A methodological refinement (Experiment 2), employing only preassociated and abstract stimuli, supported the previous findings. Although both groups of participants reported naming nearly all the stimuli to which they were exposed, post-experimental interviews again failed to reveal consistent common naming of stimuli. A final experiment investigated the effects of pre-training different stimulus naming strategies on class formation. Prior to match-to-sample training, one group of participants was trained to name stimuli individually, and another group was taught to use class-names. Although the latter group experienced more difficulty in learning stimulus names, criterion match-to-sample training took longer for the first group. Smith et al. (1996) have further reported that class-discordant associations trained orally between stimulus names prior to testing for emergent relations largely superseded previous match-to-sample baseline training, suggesting that participants' names for individual stimuli may play a role in the formation of equivalence classes, but not in their maintenance.

2.5 SUMMARY

Research leading to Sidman and Tailby's (1982) mathematical formulation of equivalence expanded the scope of the paradigm beyond reading comprehension, and the generality of their account has subsequently been validated by demonstrations of the heterogeneity of stimuli and participant populations to which it can be applied. Empirical enquiry into the formation of equivalence classes has been undertaken with regard to the number of stimuli of which those classes can be composed, and the means by which those classes can be expanded. The structure of equivalence classes has been explored in terms of nodal distance between stimuli, and demonstrations of the generalisation of equivalence classes to novel stimuli has suggested to some researchers a behavioural model for the formation of natural categories. Research into the emergence of contextually controlled equivalence classes and those composed of elements of complex stimuli have supported the theory's ecological validity, and enquiry into the delayed emergence of equivalence and its retention has further increased understanding of the temporal parameters of equivalence class formation. The theoretical revisions of Sidman's (1994) account of equivalence, and the proposal of two alternative accounts of emergent behaviour (e.g., Hayes & Hayes, 1992; Horne & Lowe, 1996) have rendered investigation of the potentially functional relationships between verbal behaviour and equivalence class formation a primary research goal in recent years. Methodological and procedural issues relating to the conduct of that research are reviewed in Chapter Three.

3. METHODOLOGICAL REVIEW

This chapter aims firstly to provide an overview of match-to-sample as the methodology of choice for the investigation of equivalence phenomena (Section 3.1), and secondly to review procedural variants of that paradigm that have also informed equivalence research (Section 3.2). A review of the methodological considerations inherent in equivalence training and testing is provided in Section 3.3, and considerations relating to both participants' and experimenters' verbal behaviour during equivalence experimentation are discussed in Section 3.4.

3.1 MATCH-TO-SAMPLE PROCEDURES

Match-to-sample is a training and testing methodology with a long and profitable history within the experimental analysis of behaviour (e.g., Sidman, 1960/1988), and owing to its simplicity, versatility, and ease of use, it has from the outset formed an empirical basis for the overwhelming majority of equivalence studies (e.g., Sidman, 1971; Sidman & Cresson, 1973; Spradlin et al., 1973). Although presentations of stimuli on paper, or by mechanical apparatus, still occur (e.g., Duarte, Eikeseth, Rosales-Ruiz, & Baer, 1998; Eikeseth, Rosales-Ruiz, Duarte, & Baer, 1997; Lazar & Kotlarchyk, 1986, MacDonald et al., 1986; Mandell & Sheen, 1994; Smeets & Barnes, 1995; Smeets, Barnes, & Roche, 1997), the possibility of experimenter effects in such procedures, and the general availability of computer systems, has led most researchers to employ computerised stimulus presentation in recent years.

A typical example of computerised match-to-sample presentation might occur as follows: A participant sits in front of a computer monitor, on which is initially presented a "sample" stimulus (for example, a red or a green square). Around this are presented three other stimuli, "comparisons" (for example, horizontal, diagonal, and vertical lines), of which the participant must chose only one using a mouse or touchscreen. During training, selection of one of these comparisons will be reinforced (i.e., a reward will be given consequent to that choice, or feedback given as to its correctness), and selection of either of the other comparison stimuli will be extinguished (i.e., reward will not be given, or feedback given that the choice was incorrect). During extinction testing, the same procedure will be employed, except that no reinforcement or punishment will be given. The relation between the sample stimulus and its correct comparison can therefore be described as one of "*if*... *then*...". For example, "*if* the sample is a red square, *then* the horizontal line is the correct comparison", and "*if* the sample is a green square, *then* the vertical line is the correct comparison". The comparison choices thus generated are referred to as "conditional" discriminations, because the selection of a specific comparison is conditional upon the presence of a specific sample. A number of procedural variations on the typical match-to-sample training and testing methodology thus described have been employed in equivalence research and are reviewed below (see Section 3.2).

The example of a match-to-sample trial presented above is an instance of what is often referred to as "simultaneous" match-to-sample (Green & Saunders, 1998), in that the sample stimulus remains in view during presentation of the correct and incorrect comparison stimuli. Some researchers have, however, additionally required an "observing" response (i.e., a selection of the sample) to be made by participants prior to comparison presentation, to promote prior orientation to the sample stimulus (e.g., Stromer & Stromer, 1990a). A common variant of the match-to-sample procedure is often referred to as "delayed" match-to-sample, in that a period of time is allowed to elapse between the removal of a previously presented sample stimulus (usually subsequent to an observing response) and the presentation of comparison stimuli (e.g., Bonta & Waters, 1981; Constantine & Sidman, 1975; Stromer, McIlvane, Dube, & Mackay, 1993). A variant of this technique has been termed the "zero delay" procedure, in that comparison stimuli are presented immediately consequent to removal of the sample stimulus (e.g., Maguire et al., 1995; Stromer & Mackay, 1993). Both of these procedures have been employed within arbitrary match-to-sample preparations, and further used to demonstrate constructed response identity-matching (see Section 3.2), and equivalence between elements of multi-element stimuli.

Another match-to-sample variant involving temporal delay between presentation of stimuli has been referred to as the "delayed-cue" procedure (Sidman, 1977; Touchette, 1971). This procedure typically involves presentation of an auditory sample stimulus (e.g., a spoken word), repeated every few seconds, in the presence of a number of visual comparison stimuli. During initial trials, an additional visual sample stimulus, identical to the correct comparison (i.e., the "delayed-cue"), also remains in view during auditory stimulus presentations. Across subsequent trials, however, presentation of the visual sample is separated by incrementally greater intervals of time from presentation of the auditory sample, until participants' correct comparison selections are reliably controlled solely by the auditory sample. The delayed-cue procedure has proved particularly effective in establishing equivalence among severely learning-disabled participants (e.g., Glat et al., 1994).

Although visual comparison stimuli are usually presented simultaneously in arrays whose spatial locations vary pseudo-randomly from trial to trial, the "successive" procedure has also been employed to establish equivalence classes composed entirely of auditory stimuli (e.g., Dube et al., 1993). Using this procedure, a response to a visual stimulus whose location is invariant results in presentation of the next in a sequence of auditory sample and comparison stimuli. Another variant of match-to-sample training has been termed "errorless training" (Bentall, et al., 1993, 1999; Gast et al., 1979; Holth & Arntzen, 1998). The most commonly used version of this technique initially presents only the correct comparison stimulus and its sample. Consequent to successful selection of the former stimulus, incorrect comparisons are gradually added to the stimulus array presented on each trial, dependent upon participants' continued selection of the correct comparison. The number of incorrect stimuli presented is increased thus until the number of incorrect comparisons finally to be employed is attained.

3.1.1 Training

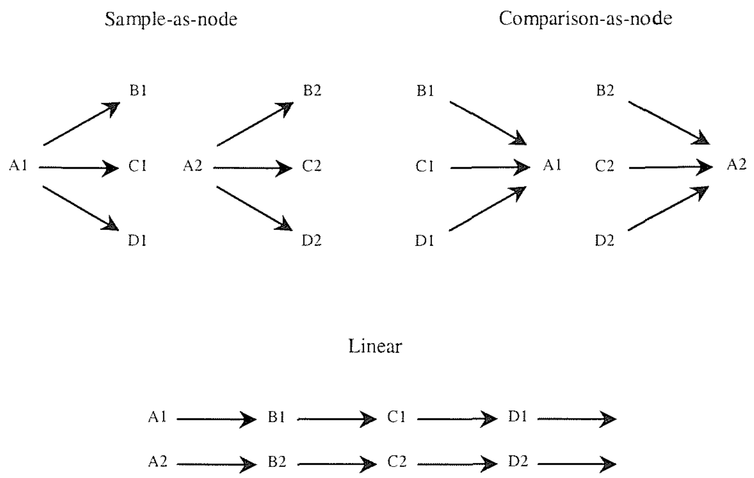

Figure 3. Three training structures commonly employed to establish stimulus equivalence. Arrows point from samples to comparisons, and indicate trained relations between stimuli in two potential four-member equivalence classes (i.e., A1, B1, C1, D1 and A2, B2, C2, D2).

A major procedural consideration within the equivalence literature has been that of the match-to-sample training structures employed during establishment of baseline relations. Three major variants can be identified (see Figure 3) that are often referred to respectively as "one-to-many", "many-to-one", and "linear" training (Green & Saunders, 1998; Saunders, Saunders, Williams, & Spradlin, 1993; Sidman, 1994). Assessment of the effects of these training structures has often been closely linked with enquiry into the effects nodal distance on equivalence class formation (cf. Fields et al., 1984).

As Saunders et al. (1993) have pointed out, one-to-many training--also sometimes referred to as "single-sample, multiple-comparison" (Saunders et al., 1988) or "sample as node" training (Fields et al., 1984)--has been used within a large number of the equivalence studies that have employed human participants (e.g., Barnes et al., 1990; Cowley et al., 1990; Devany et al., 1986; Dugdale & Lowe, 1990; Hayes et al., 1991; Lazar & Kotlarchyk, 1986; Pilgrim & Galizio, 1995; Sidman et al., 1985, 1986; Wulfert, Dougher, & Greenway, 1991). Using this technique, a single sample stimulus controls selection of several comparison stimuli, and the sample thus serves as a nodal stimulus. A second technique, whose use has been less frequently reported within the equivalence literature (e.g., Green et al., 1991; Saunders et al., 1988; Sigurdardottir et al., 1990; Spradlin et al., 1973; Spradlin & Saunders, 1986), is many-to-one training, also sometimes referred to as "multiple-sample, single-comparison" (Saunders et al., 1988) or "comparison as node" training (Fields et al., 1984). Using this technique--which is in many respects the inverse of one-to-many training--several sample stimuli control selection of a single comparison stimulus, and the comparison thus serves as a nodal stimulus. A third training technique is that of "linear" or, as it is sometimes also termed, "internal node" training (e.g., Dube et al., 1989; Fields et al., 1990; Kennedy, 1991; Lazar et al., 1984; Lynch & Green, 1991). Using this structure, a linear series of conditional discriminations is established in which each internal stimulus (i.e., all stimuli except the first and last stimuli in a given class) is a nodal stimulus.

Although these training structures have often been assumed to be equally likely to promote the emergence of equivalence (cf. Saunders et al., 1993), a number of studies have nonetheless employed a variety of stimulus materials and participants to test whether equivalence class formation is related to baseline training structure (e.g., Arntzen & Holth, 1997; Saunders et al., 1999, 1988;

Smeets, Leader, & Barnes, 1997; Spradlin & Saunders, 1986; Spradlin et al., 1992; Urcuioli & Zentall, 1993). Spradlin and Saunders (1986) and Saunders et al. (1993, 1988), for example, have demonstrated that equivalence is more likely to emerge as a result of many-to-one training, using abstract visual stimuli presented to intellectually-disabled participants. As the latter authors have noted, however, such effects may be diminished among mainstream participant populations. Although such findings have subsequently been supported by Saunders et al. (1999) using mainstream children presented with abstract visual stimuli, Arntzen and Holth (1997) have nevertheless reported that equivalence may be more likely to emerge as a result of one-to-many training among mainstream adults presented with abstract visual stimuli. Research into the possible relationships between respondent conditioning and the training protocols outlined above is presented in Section 3.2.1.

3.1.2 Testing

Another major procedural consideration within the equivalence literature has been that of the testing structures employed during assessment of equivalence class formation. Although equivalence studies have employed a variety of procedures to test for such emergent relations, three major variants can be identified, that are often referred to as "complex to simple", "simple to complex", and "simultaneous" protocols.

As Adams et al. (1993a) and Fields et al. (1997) have noted, the complex-to-simple protocol denotes an experimental preparation in which trials of transitivity and equivalence are presented in extinction immediately subsequent to establishment of baseline relations: If the emergence of equivalence is not observed during these trials, symmetric relations alone are tested. If symmetric relations can be demonstrated to have emerged, relations of transitivity and equivalence are then tested again. A number of equivalence studies have successfully employed this technique using a variety of stimuli and participants (e.g., Bush et al., 1989; Devany et al., 1986; Fields et al., 1992, 1990; Kennedy & Laitinen, 1988; Lazar et al., 1984; Saunders et al., 1988; Sidman, 1971; Sidman & Cresson, 1973; Sidman et al., 1985; Spradlin et al., 1973). As Adams et al. (1993a) and Fields et al. (1997) have further noted, another preparation that has less frequently used is that referred to as the simple-to-complex protocol. Using this technique, all component baseline relations are established individually, each followed immediately by tests of the symmetry of those specific relations. Subsequently, tests of all other emergent relations are presented (e.g., Adams et al., 1993a; Fields et al., 1991; Lynch & Cuvo, 1995). As Fields et al. (1997) have noted, although both techniques have reliably produced equivalence across a variety of human participant populations, the latter technique has usually produced equivalence subsequent to presentation of fewer test trials than the former.

A third procedure, referred to as the "simultaneous" protocol, has also formed the basis of equivalence research (Fields et al., 1997). Using this technique, all baseline conditional discriminations are introduced in a single block of randomly presented trials. Training continues in this way until all baseline relations have been established and, subsequently, testing blocks composed of trials involving all relations of symmetry, transitivity, and equivalence are presented in random order. As these researchers have noted, however, the simultaneous protocol is often less effective in producing stimulus equivalence, although successful demonstrations of equivalence have been accomplished as a result of such training (e.g., Buffington, Fields, & Adams, 1997; Kennedy et al., 1994).

A further methodological issue relating to the testing of emergent relations has been highlighted by Green and Saunders (1998), who have noted that many early equivalence studies (e.g., Sidman et al., 1985; Spradlin et al., 1973) presented trials of emergent relations to participants as probes within blocks of reinforcement trials involving previously established baseline relations. Subsequent research, however, has often tested for emergent relations in discrete trial blocks containing no trials of baseline relations. As Green and Saunders (1998) have pointed out, however, this procedure introduces a potential confound: Failure to demonstrate equivalence may actually result from a failure in baseline maintenance (cf. Stikeleather & Sidman, 1990). Although it is possible to reduce the occurrence of this artefact by the assessment of baseline maintenance immediately prior to and subsequent to emergent testing blocks, little empirical evidence is available as to the efficacy of testing for emergent relations using probe as opposed to massed trial presentations. Research into the relations between respondent conditioning and the testing protocols outlined above is presented in Section 3.2.1.

3.2 ADDITIONAL PROCEDURES

As noted above, the overwhelming majority of equivalence studies have employed match-to-sample procedures, both to establish baseline relations and to test for the emergent relations of equivalence. Use of a variety of additional experimental preparations has nevertheless been reported, and the most important of these are reviewed in sections 3.2.1 to 3.2.6.

3.2.1 Respondent Conditioning

Rehfeldt and Hayes (1998a) have suggested that the distinction commonly drawn between operant and respondent conditioning may represent an impediment to a fuller understanding of stimulus equivalence. As these researchers have pointed out, nearly all equivalence studies have assayed equivalence phenomena from the perspective of operant conditioning. Nevertheless, a number of studies have directly set out to investigate the role of respondent conditioning in equivalence class formation.

Leader, Barnes, and Smeets (1997), for example, employed a "respondent-type" training technique to demonstrate that the exposure to differential reinforcement characteristic of the majority of equivalence studies is not a requirement for equivalence. During establishment of baseline relations, adult participants were simply exposed to pairs of visual nonsense-syllables, presented sequentially and separated by varying inter- and intra-pair temporal delays. Participants were not required to respond on these trials, but simply to attend to the stimuli presented. Match-to-sample testing in extinction subsequently revealed that the stimuli presented had become related by equivalence, and also that equivalence class formation occurred most reliably among those participants exposed to longer intra-pair delays relative to within-pair delays. The conclusion was thus reached that temporal contiguity between stimuli alone may suffice to establish equivalence.

In a replication and extension of these findings (Smeets et al., 1997), mainstream adults and children were employed as participants to assess the effects of respondent conditioning on equivalence class formation, using a variety of different training (i.e., many-to-one, one-to-many, and linear) and testing (i.e., simultaneous and simple-to-complex) techniques. The results again indicated the efficacy of stimulus pairing techniques in the promotion of equivalence and additionally indicated that, among adult participants, symmetry and equivalence emerged most rapidly as a result of training and testing using simultaneous and many-to-one protocols. Among children, however, the results indicated that the simultaneous protocol was ineffective in promoting equivalence, and that all testing protocols resulted in similar emergent performances when simple-to-complex training had been employed.

Roche and Barnes (1997) have further provided evidence of the transformation of respondently conditioned stimulus functions in accordance with arbitrarily applicable relations. Adult male participants were presented with a sexual film clip (e.g., of an act of felatio or cunnilingus) paired with a visual nonsense syllable (e.g., JOM or ROG), and an emotionally neutral film clip (e.g., from a geographic documentary) paired with another nonsense syllable. Subsequent match-to-sample training established equivalence relations between the former nonsense syllables and sets of novel nonsense syllables. Measures of skin conductance indicated that emotional responses had transferred to the nonsense syllables previously paired with sexual film clips, and that a transformation of eliciting stimulus functions had taken place via the establishment of equivalence classes. Other researchers (Augustson & Dougher, 1997; Dougher, Augustson, Markham, Greenway, & Wulfert, 1994) have likewise reported the transfer of eliciting, extinction, and avoidance evoking functions through equivalence classes using respondent conditioning techniques.

Although usually discussed within a framework of operant conditioning, some researchers (e.g., Rehfeldt & Hayes, 1998a; Stromer et al., 1993) have pointed out that the presentation of multi-element stimuli employed in equivalence research has potentially provided the conditions of temporal contiguity between stimuli that have been demonstrated to establish equivalence within the respondent paradigms outlined above.

3.2.2 Simple Discrimination Training

A simple discrimination typically describes the consistent differential selection of one of two or more stimuli, the selection of which is correlated with reinforcement. For example if, given stimuli A1 and A2, selection of A1 (S+) were to be reinforced, selection of A2 (S-) would be extinguished. The first empirical demonstration of the facilitative effects of training component simple discriminations on the emergence of bidirectional relations was provided by Saunders and Spradlin (1989), who presented abstract visual forms to two intellectually-disabled adult participants with extensive histories of failure on arbitrary match-to-sample tasks. Both participants were trained with the component simple discriminations involved in two choice conditional discriminations (i.e., successive discriminations between sample stimuli and simultaneous discriminations between comparison stimuli). Subsequent to such training, both participants demonstrated the emergence of symmetric relations between sample and comparison stimuli during match-to-sample testing.

More recent research involving simple discrimination training has employed both mainstream adults and children as participants. Schenk (1995), for instance, employed children as participants to demonstrate that no-reinforcement presentation of multi-element visual stimuli using a simple discrimination procedure can result in the emergence of untrained simple and conditional relations between the elements of those stimuli. Smeets and Barnes (1997) have further examined whether trained and derived simple discriminations can lead to conditional relations between stimuli serving the same and opposite (i.e., S+ or S-) functions, using both adults and children as participants. Although the children's performance provided only inconclusive findings, that of the adult participants indicated that they had matched all paired and conditionally related stimuli to one another, thus extending the previous findings of Smeets, Barnes, and Luciano (1996), who had employed only children as participants. The results of both studies led Smeets and Barnes (1997) to conclude that temporal contiguity alone may provide an explanation for the findings of studies that have employed "outcome-specific" training procedures (see Section 3.2.4), and for the emergence of novel conditional discriminations more generally.

3.2.3 Observational Learning

In addition to the studies cited above, the emergence of stimulus equivalence has been reported as a specific result of both observational learning and direct training, using intellectually-disabled adults as participants (MacDonald et al., 1986). In this study, participants were directly trained with one conditional discrimination involving stimuli from each of two potential equivalence classes, and then given the opportunity to observe another participant being trained with similar conditional discriminations involving other stimuli from the same potential equivalence classes. Each observed relation presented one stimulus that had previously been presented during direct training. Subsequent testing in extinction revealed that all stimuli that had been related by direct and observed training had also become related by equivalence. Both MacDonald et al. (1986) and Stoddard and McIlvane (1986) have noted the potential educational applications for the establishment of emergent behavioural repertoires through observational learning techniques.

3.2.4 Outcome-specific Procedures

As mentioned above, some equivalence studies have employed outcome-specific training procedures (i.e., procedures in which a number of reinforcing stimuli are employed differentially to reinforce specific conditional discriminations) to establish the prerequisite relations of equivalence among intellectually-disabled participants, using a variety of arbitrary auditory and visual stimuli (e.g., Dube et al., 1987, 1989; Schenk, 1995).

Dube et al. (1987), for instance, demonstrated that reinforcers can participate in equivalence classes, and also that novel stimuli can become class members through relations with the former stimuli. Dube et al. (1989) provided a replication and extension of this finding using a similar experimental procedure and similar participants, but other reinforcing stimuli (i.e., a variety of foodstuffs). In both of these studies, identity matching (see Section 3.3.4) or arbitrary matching performances were established using differing reinforcers whose presentation was specific to the conditional discriminations that preceded them on training trials. Extinction testing subsequently revealed that training had established bidirectional relations between the differing consequential stimuli and the comparisons whose selection they had previously reinforced.

As Sidman (1994) has observed, outcome-specific reinforcement procedures make it possible to test whether stimuli that have previously served to reinforce conditional discriminations can subsequently serve as samples and comparisons. Smeets and Barnes (1997) have further suggested that temporal contiguity alone may provide sufficient explanation for the findings of the studies presented above, and thus also account for observations that conditional and reinforcing stimuli can become related by equivalence.

3.2.5 Sequence Training

Lazar (1977) was the first researcher to provide an empirical demonstration that adult participants, trained with sequences of abstract visual stimuli, can subsequently demonstrate the prerequisite relations of equivalence between the stimuli in those sequences. On the basis of these findings, it was proposed that stimulus equivalence may underlie humans' learning of simple grammar and syntax. A replication and extension of these findings was provided by Lazar and Kotlarchyk (1986), who demonstrated the contextual control of sequence-classes composed of abstract visual, using children as participants. A number of subsequent studies have provided similar findings.

Wulfert and Hayes (1988), for example, established two four-member equivalence classes composed of abstract forms among adult participants. When participants were trained to select one stimulus from each class in a given order, subsequent testing revealed that those participants consistently ordered all other members of both equivalence classes in the absence of further training. Both this ordering, and subsequently established conditional sequential ordering, was then observed to have transferred to all members of both equivalence classes. Potential similarities were again noted by these researchers between the formation of sequence-classes of equivalent stimuli and learning of syntactical structures.

The potential of sequence training as a tool for the establishment of basic reading repertoires has further been indicated by Stromer and Mackay (1992), who demonstrated that the spelling performance of intellectually-disabled participants can be improved using delayed constructed-response identity matching procedures (i.e., through establishment of relations between easily nameable pictures and their written names, and participants' sequential ordering of the written letters composing those stimulus names). Maydak, Stromer, Mackay, and Stoddard (1995) have also provided evidence of the utility of sequence training in producing numeric relations among intellectually-disabled participants. Green, Sigurdardottir, and Saunders (1991) have further extended these findings, and those of Lazar and Kotlarchyk (1986), by demonstrating the transfer of ordinal functions through equivalence classes using abstract forms presented to adult participants.

Finally, Green, Stromer, and Mackay (1993) have proposed an analysis of emergent performances derived from the contingencies that establish the production of stimulus sequences. As these researchers have noted, neither simple chaining, nor within-sequence conditional control can account for the production of the novel sequences observed in the experiments outlined above. Extending Sidman and Tailby's (1982) mathematical analysis of stimulus equivalence, therefore, Green et al. (1993) proposed that stimulus sequences can be understood in terms of the definitional relational properties of an order relation (i.e., asymmetry, transitivity, and connectedness). Unfortunately, little research has resulted from the theoretical issues raised in this paper.

3.2.6 A Precursor to the Relational Evaluation Procedure

As Cullinan, Barnes, and Smeets (1997) have noted, previous research (Leader et al., 1996) has documented the use of a respondent-type procedure that has proved effective in the generation of stimulus equivalence. In that experiment, baseline establishment was composed solely of the sequential presentation of pairs of visual stimuli to which participants were not required to respond, but simply to attend. Subsequent match-to-sample testing in the absence of reinforcement demonstrated that the stimuli presented had become related by equivalence.

In an attempted extension of these findings, Cullinan et al. (1997) reported the use of a further non-match-to-sample procedure by which the prerequisite relations of stimulus equivalence may be both trained and tested (the "a precursor to the Relational Evaluation Procedure" technique). Again employing adult participants presented with visual nonsense-syllables, a single sample stimulus was presented on every trial, followed by presentation of either a positive or a negative comparison. Using a "go/no go" procedure, participants were required to respond to correct comparisons, but not to respond to incorrect comparisons: Responses to correct and incorrect comparisons were respectively reinforced and punished. The performance of participants exposed to this technique was compared to that of similar participants presented with the same stimuli using a match-to-sample procedure. The findings indicated that participants exposed to training and testing using the "a precursor to the Relational Evaluation Procedure" technique were less likely to demonstrate equivalence than those exposed to match-to-sample training and testing. Participants who had received exposure to match-to-sample trials prior to exposure to the "a precursor to the Relational Evaluation Procedure" technique were also found to be more likely to demonstrate equivalence than those participants who had not been so exposed. Although Cullinan et al. (1997, p. 121) have noted that their findings "may contribute to a fuller understanding of those variables most relevant to producing equivalence responding", no other studies have, as yet, utilised the "a precursor to the Relational Evaluation Procedure" technique further to push back the boundaries of equivalence research.

3.3 METHODOLOGICAL CONSIDERATIONS

Whereas the aim of sections 3.1 and 3.2 was to provide a review of the match-to-sample procedures and procedural variants typically employed in equivalence research, Section 3.3 aims to review the methodological considerations inherent in the use of those procedures.

3.3.1 Designations and Criteria

The stimulus classes that form the basis of equivalence research, and the stimuli of which those classes are composed, are typically referred to by alphanumeric designations. Although the choice of designations employed is necessarily arbitrary, for ease of reference, and to facilitate comparison between studies, a descriptive convention has been consistently maintained within the equivalence literature: Stimuli are referred to by alphabetic designations, and stimulus classes by numeric designations. For example, three-member sets of stimuli might be labelled "A", "B", and "C", and the three potential equivalence classes to which they belong might be labelled "1", "2", and "3". Hence, within a linear training procedure, conditional relations would be established between "A1" and "B1", and between "B1" and "C1". Similar conditional relations would subsequently also be established between the "A", "B", and "C" stimuli participating in potential equivalence classes "2" and "3".

Regarding the performance criteria employed during training of baseline relations, a very high level of accurate responding is required. As Stikeleather and Sidman (1990, p. 2) have observed, "insufficiently rigorous criteria for the acquisition of [baseline] conditional discriminations can cause a failure to establish the necessary prerequisites [of equivalence]--a failure that is likely to go unrecognized". As these and other researchers (e.g., Green & Saunders, 1998) have also observed, two choice match-to-sample preparations afford the potential that performance at as high a level of accuracy as 75% can nevertheless result from chance factors. Many researchers have therefore recommended acceptance of participants' mastery of baseline relations at a minimum of 90% or, preferably, 100% accuracy (Sidman, 1994). The number of baseline trials across which participants must demonstrate such accuracy remains arbitrary, however, but represents an issue that must be decided prior to research (Green & Saunders, 1998).

As noted above, many recent equivalence studies have assessed the emergence of equivalence through the presentation of blocks of massed emergent trials. Use of such procedures, however, precludes simultaneous assessment of baseline maintenance, as mentioned previously. It is therefore usually held to be necessary to assess baseline maintenance in extinction prior to emergent testing and, if equivalence has not been shown initially to have emerged, prior to any subsequent re-testing. The criterion at which equivalence is accepted to have been demonstrated is also an arbitrary decision. Again, most researchers have favoured high levels of class-consistent responding, usually reported at between 90% and 100% accuracy (Sidman, 1994).

3.3.2 Reinforcers and Comparisons

The use of a variety of consequential stimuli has been reported within the equivalence literature. Although the choice of stimuli to be so employed is usually governed by the participant populations to which they are presented, and the experimental preparations within which they are employed, stimuli that have been observed to function as reinforcers have included tokens, money, points, brief verbal praise, musical chimes, and foodstuffs. Likewise, punishing consequences reported within the literature have included brief verbal reprimands, buzzing noises, or the deduction of previously earned points, tokens, or money. Consequential stimuli are usually presented immediately subsequent to comparison selection during training of baseline relations. Selection of correct comparisons results in presentation of a reinforcer, and selection of any other comparison stimulus usually results in presentation of a punisher. Some equivalence studies have, however, employed simple extinction procedures (i.e., non-presentation of a reinforcer, or a brief "time out") in lieu of direct punishment. During assessment of baseline maintenance in extinction and testing of emergent relations, no consequential stimuli are presented subsequent to selection of any comparison.

With rare exceptions (e.g., Bentall et al., 1993; Gast et al., 1979; Saunders et al., 1988; Sigurdardottir et al., 1990; Smith et al., 1996), equivalence studies have, throughout training and testing, employed as incorrect comparisons all members of stimulus classes sharing the same alphabetic designation as the correct comparison. For example, given four three member potential equivalence classes composed of stimuli "A1", "B1", "C1"; "A2", "B2", "C2"; "A3", "B3", "C3"; "A4", "B4", "C4", baseline training might initially present "A1" as sample, with "B1" as its correct comparison. Stimuli "B2", "B3", and "B4" would therefore be presented as the incorrect comparisons presented on this trial. During subsequent emergent testing, "B1" would be presented as sample, and "A1" as the comparison whose selection would indicate the symmetry of that trained relation: The incorrect comparisons on this trial would be "A2", "A3", and "A4". Although there is again no *a priori* reason why this convention should be adhered to, ease of reference and facilitation of comparison between studies have nonetheless fostered its maintenance.

A further issue relating to the presentation of comparison stimuli is their spatial location within stimulus arrays. As a number of researchers (e.g., Sidman, 1994; Stikeleather & Sidman, 1990) have pointed out, consistent spatio-temporal patterns in the presentation of such stimuli can result in artefactual findings. Equivalence research has therefore favoured the randomised, or quasi-randomised, allocation of comparison stimuli to various locations within the arrays in which they are presented.

3.3.3 Inter-trial Intervals

An inter-trial interval is the time that is allowed to elapse between the cessation of one trial presentation and the onset of the following trial presentation. Unfortunately, because only a minority of researchers have reported the inter-trial intervals employed during their training and testing procedures, little data is available with regard to the potential relationships between the length of such intervals and equivalence class formation. From the data available, however, various researchers (e.g., Sidman, 1994; Stromer & Mackay, 1996) have noted that there may be possible correlations between participants' naming of stimuli and the inter-trial intervals employed in equivalence research.

Sidman (1994) for example, has reported the use of inter-trial intervals of .67 s in research from his laboratory, and further noted that the rapid presentation of consecutive trials implied may preclude the emission of verbal behaviour by participants exposed to such procedures. Maydak et al. (1995) have reported use of inter-trial intervals of 3-s in their research using participants with intellectual disabilities, and Kennedy (1991) has reported inter-trial intervals of approximately 5-s. Other researchers (e.g., Pilgrim, Galizio, & Chambers, 1995) have used inter-trial intervals as great as 20-s, but nevertheless reported that the children employed as participants in their research exhibited neither common nor intraverbal naming of the abstract objects presented to them as stimuli.

3.3.4 Reflexivity

As described above, reflexivity is one of the three defining relations of equivalence. It does however possess a unique property that differentiates it from both symmetry and transitivity. Whereas both latter terms denote relations between two or more different stimuli (for example, *a***R***b* and *a***R***c*, and *a***R***c* and *c***R***a*, respectively), reflexivity describes a relation between a stimulus and itself (i.e., *a***R***a*, *b***R***b*, and *c***R***c*). Reflexivity has therefore been described as "identity matching" (Sidman et al., 1982), in that it requires the selection of a comparison stimulus conditionally upon presentation of an identical stimulus as sample (i.e., as a result of physical sameness).

Early equivalence research (e.g., Sidman, 1971; Sidman & Cresson, 1973) often presented tests for reflexivity prior to baseline training. Because reflexivity represents one of the emergent relations of stimulus equivalence, however, assessment of identity matching prior to baseline training is not easily justified, because the results observed can only be determined by factors other than those resulting from the establishment of baseline relations (Carrigan & Sidman, 1992; Johnson & Sidman, 1993; Saunders & Green, 1992). Bush et al. (1989) were the first researchers to accept reflexivity as a given for adult participants, however, and tests of reflexivity were not presented in their research. The majority of subsequent equivalence studies using similar participant populations have been constructed likewise.

3.3.5 Measures

As Sidman (1986, p. 329) has pointed out, "for behaviour analysts, the primary object of observation and measurement, behavior itself, is usually conceptualized as a *response* [emphasis in original]". Equivalence research has maintained this tradition through its emphasis on the assessment of participants' class-consistent and non-class consistent responses on trials of both baseline and emergent relations. Although these responses have almost invariably provided the primary measure of baseline establishment, maintenance, and equivalence class formation, additional measures have been employed in equivalence research--and response latencies have proved one of the most informative within a variety of experimental preparations.

A response latency can be defined as the amount of time that elapses between the presentation of comparison stimuli and a participant's selection of one of those stimuli (i.e., usually, for the purpose of measurement, the correct stimulus). Participants' observing response latencies can also be measured, but have seldom been assessed in equivalence research. Wulfert and Hayes (1988) were among the first researchers to employ participants' response latencies as a measure of the effects of nodal distance on equivalence class formation and, subsequently, many other researchers have evaluated their findings likewise (e.g., Bentall et al., 1993, 1999; Dickins et al., 1993; Fields et al., 1990; Kennedy, 1991; Kennedy et al., 1994). As Spencer and Chase (1996) have pointed out, participants' latencies can provide a more sensitive measure of performance than accuracy alone.

Use of other response measures has been reported by a small number of other researchers. These have included assessment of skin conductance (e.g., Augustson & Dougher, 1997; Dougher et al., 1994; Roche & Barnes, 1997), stimulus class ratings (Lane et al., 1999), and "yes/no" and "can't answer" response options (Duarte et al., 1998; Eikeseth et al., 1997). Analysis of participants' verbal behaviour has also been employed in a number of studies (see sections 3.4.2 and 3.4.3).

3.4 VERBAL CONSIDERATIONS

From the outset, equivalence research and the study of verbal behaviour have been closely related. It is unsurprising, therefore, that considerations relating to the verbal behaviour of both participants and researchers during equivalence research have received attention within the experimental literature. What is perhaps surprising, however, is the relatively small number of studies that have specifically addressed the issues involved. The aim of Section 3.4.1 is to provide a review of studies that have addressed the role of the instructions in equivalence research, and the theoretical questions they have raised. The aim of sections 3.4.2 and 3.4.3 is to provide a review of methods by which participants' verbal behaviour during experimentation has been directly assessed, and various criticisms offered regarding those methods.

3.4.1 Instructions

Although the term "instruction" can be used to describe the function of a sample stimulus that serves as a consistent cue for correct comparison selection in match-to-sample procedures (Lowenkron & Colvin, 1995; Sidman, 1994), the term is almost always employed to describe either spoken or written verbal directives to participants with regard to their conduct during experimentation. A related use of the term describes the induction of participants' naming of experimental stimuli during equivalence research.

Regarding the second and arguably the most far-reaching usage, the majority of equivalence studies have sought to minimise instruction to participants (for example, Bush et al. (1989) simply instructed participants to "touch it"), to avoid the contamination of purely contingency-shaped behaviour by verbal control. As Dymond and Barnes (1997) have pointed out, however, only a small number of studies have systematically examined the ways in which instructions can affect the formation of equivalence classes (Devany et al., 1986; Duarte et al., 1998; Eikeseth et al., 1997; Saunders et al., 1993; Spencer & Chase, 1996; Wulfert & Hayes, 1988). The majority of these studies have set out to investigate instructional effects with regard to specific procedural issues involved in equivalence research.

Saunders et al. (1988), for instance, compared mildly intellectually-disabled participants given experimental instructions (e.g., "when A1 appears, press B1; when A2 appears, press B2", etc.) with similar participants given no such instructions. Their findings indicated that even the latter participants demonstrated equivalence using both many-to-one and one-to-many training, leading to the conclusion that instructions are not necessary for the formation of equivalence classes. Spencer and Chase (1996), using arbitrary visual forms as stimuli, compared the performance of adult participants given minimal instructions (i.e., statements regarding the feedback with which they would be presented during training) with those of similar participants given extensive verbal descriptions of the match-to-sample contingencies of the experiment (i.e., rules indicating the experimenter-designated classes of stimuli employed) to assess equivalence class formation with regard to participants' speed and accuracy of responding. Their findings indicated no significant differences either in the accuracy with which participants in the different groups responded, or in the effects of nodal distance observed in participants' response latencies.

Eikeseth et al. (1997) examined the effects of instructions on equivalence class formation using a procedure involving paper-and-pencil training and testing. In this experiment, printed Roman letters were presented to adult participants who either had, or had not, received a detailed written description of the match-to-sample contingencies of the experiment. Participants selected comparisons by marking their choices in pencil on the sheets of paper upon which the experimental stimuli were printed. A greater proportion of participants who had previously received instructions demonstrated equivalence compared to those participants who had received no such instructions. Duarte et al. (1998) further investigated the effects of different types of instructions on equivalence class formation, in a study that employed participants and pencil-and-paper procedures similar to those described above. All participants in this study were initially given instructions regarding the specific sample-comparison pairings governing correct responding during trials of baseline relations. One group of participants was given no further instructions, but a second group was given additional "restrictive" instructions, cautioning against "going beyond" the instructions with which they had previously been presented. A third group of participants received "non-restrictive" instructions, urging consideration of the additional relationships between stimuli implicit in their original instructions. Some participants in each group were additionally provided with a "can't answer" response option on every trial. Findings indicated that the emergence of stimulus equivalence was markedly affected by the availability of a "can't answer" response option, and also by the nature of the instructions given: Participants who had received restrictive instructions were found to be less likely to demonstrate equivalence than those who had received non-restrictive instructions.

Sigurdardottir et al. (1990) examined the effects of instructions on the formation of equivalence classes generated as a result of sequence training, by comparing the performance of participants who had received instructions regarding the sequencing and match-to-sample contingencies of the experiment with that of other participants who had received no such instructions. Few differences were observed between the performance of participants in either group during establishment of sequence responding and equivalence class formation, however, although participants who had received no instructions required repeated training and testing to demonstrate equivalence. Relationships were thus suggested between instructional effects and the delayed emergence of equivalence.

Other studies have investigated the effects of instructions on the transfer of functions. Green et al. (1991), for instance, set out to assess whether the transfer of ordinal functions through equivalence classes would be affected by instructions. Surprisingly, their findings indicated that participants exposed to comprehensive instructions (i.e., instructions regarding the sequencing and match-to-sample contingencies of the experiment) demonstrated transfer of ordinal functions less readily than participants who were provided with only minimal instructions (i.e., instructions to initiate interactions with the apparatus and the contingencies). Research by Dymond and Barnes (1994, 1997) has additionally indicated that detailed verbal instructions (i.e., thorough descriptions of the experimental procedure) are unnecessary for adult participants to demonstrate the successful transfer of self-discrimination functions.

With regard to participants' experimentally-induced naming of stimuli, Eikeseth and Smith (1992) have reported that instructions to provide common names for experimental stimuli can facilitate the emergence of equivalence among autistic participants who have previously tested negative for equivalence. Another study, comparing the effects of differing instructions on equivalence class formation (Smith et al., 1996) has further suggested that class-discordant associations trained orally between the names of easily-recognisable visual stimuli prior to testing for emergent relations can largely supersede prior match-to-sample training among adult participants. These findings led to the conclusion that although participants' naming of individual stimuli may play a role in the formation of equivalence classes, such naming may not remain functional in their maintenance.

3.4.2 Verbal Reports

Used generally, the term "verbal report" can describe any written or spoken statement delivered by a participant to an experimenter, before, during, or after experimentation. As Critchfield et al. (1998, p. 436) have pointed out, "self-reports can provide information about a vast array of behavioral phenomena, many of which would be difficult to measure in other ways." And, in fact, "self-report data may provide the only practical means of observing certain forms of behavior" (Perone, 1988, p. 72). Nevertheless, behaviour analysts have traditionally remained antipathetic towards the use of verbal reports (e.g., Sidman et al., 1986) on the grounds that the data thus garnered may not accurately reflect the contingency-shaped performance ostensibly described (e.g., Shimoff, 1984, 1986). The use of intra-experimental verbal reports has been criticised for its potential prompting of participants to provide names for stimuli that would otherwise have remained unnamed (Sidman, 1992, 1994), and the use of post-experimental verbal reports has likewise been attacked because of the implication that participants may be compelled to provide names for stimuli that might otherwise not have been named, or that might have been named otherwise during experimentation (Dugdale & Lowe, 1990; Sidman, 1994; Stoddard & McIlvane, 1986). A number of behavioural researchers have, however, supplemented their research with the use a variety of self-report methods in order further to elucidate participants' verbal behaviour during equivalence training and testing.

An early example of this approach was provided by Sidman and Tailby (1982), who employed oral tests of their child participants' common naming during experimentation, on trials in which only a sample stimulus was presented. No reinforcement resulted from participants' verbal responses, only presentation of another sample stimulus. Although participants' verbalisations during prior match-to-sample training and testing were additionally analysed, neither measure indicated that participants' naming of stimuli was either necessary or sufficient for equivalence class formation. Sidman et al. (1986) employed a similar methodology to test for the use of common names among the mainstream children and intellectually-disabled adults employed in their research. Their findings indicated that during auditory-visual match-to-sample trials, half of their intellectually-disabled participants failed to apply the names of auditory stimuli to each visual stimulus in the same class and, additionally, that when classes of purely visual stimuli were presented, all but one child failed to apply a common name to those stimuli. Because all participants nonetheless demonstrated equivalence, the conclusion was again reached that common naming was neither necessary nor sufficient for the formation of equivalence classes.

Bentall et al. (1993) have further employed post-experimental interviews and post-experimental written tests of naming to assess whether adult participants name visual stimuli presented during equivalence research. Findings indicated strong correlations between participants' self-reports and their performance during match-to-sample training and testing. Further research has employed a similar post-experimental verbal report methodology to assay participants' potential verbalisations on match-to-sample tasks. Trigo, Rafael, and Moreno (1995), for instance, trained adult participants on a matching task using geometrical figures, and subsequently asked them to describe the "key" to their performance. Generalisation testing revealed that only those participants who had stated a general rule were able to satisfy the task requirements.

The most widely employed method of obtaining and analysing participants' verbalisations during equivalence research, however, has been through the collection and analysis of concurrent verbal reports. These are usually assessed under the rubric of "protocol analysis" (see Section 3.4.3): Applications of this procedure to the experimental analysis of human behaviour are reviewed below.

3.4.3 Protocol Analysis

As a tool for the evaluation of human "cognitive processes" processes, protocol analysis (Ericsson & Simon, 1980, 1993, 1998) has, for some years, enjoyed the esteem of many cognitive researchers (e.g., Crutcher, 1994; Mack, 1985; Magliano, 1996; Payne, 1994; Svenson, 1985; Venkatesan, 1986; Wilson, 1994), although others have remained sceptical (e.g., Broadbent, 1986; Duncan, 1985; Laffal, 1985; Solomon, 1995). In recent years, a number of behaviour analysts have suggested protocol analysis as a means of investigating participants' verbal behaviour during experimentation (e.g., Austin & Delaney, 1998; Critchfield, Tucker, & Vuchinich, 1998; Hayes, 1986; Hayes, White, & Bissett, 1998; Lane & Critchfield, 1996), and a number of empirical studies have implemented this suggestion with regard to the formation of equivalence classes (e.g., Potter, Huber, & Michael, 1997; Rehfeldt et al., 1998; Wulfert et al., 1991, 1994). Although Ericsson and Simon (1980, 1993, 1998) have laid down stringent criteria for the application of their technique, "there currently exist no standards for self-report methods specific to the experimental analysis of behaviour" (Critchfield et al., 1998, p. 436) and hence, a variety of experimental preparations have been documented within the behavioural literature.

The first equivalence study to employ protocol analysis was reported by Wulfert et al. (1991), who set out to investigate individual differences in the emergence of stimulus equivalence using visual stimuli presented to adult participants using a "think-aloud" procedure (Ericsson & Simon, 1980, 1993). The results observed led to the conclusion that participants who described stimulus compounds (e.g., "together the stimuli look like a house") were less likely to demonstrate equivalence than those who described relations between stimuli (e.g., "circle goes with the open triangle"). These findings were subsequently extended by Wulfert et al. (1994) who, using similar stimuli and participants, reported that the verbal behaviour of participants who failed to demonstrate the formation of third-order equivalence classes indicated that the requisite control by second-order conditional stimuli had not been established by the contingencies of the experiment. When this control was established through verbal instruction, those participants who had previously failed so to do demonstrated the formation of third-order classes, thus indicating the utility of protocol analysis as a measure of stimulus control during equivalence research.

A further demonstration of the utility of protocol analysis in the investigation of stimulus equivalence has been provided by Lane and Critchfield (1996), who again employed arbitrary visual forms presented to adult participants using a think-aloud procedure. Participants were initially trained to categorise their responses on a conditional discrimination task as either correct or incorrect. Subsequent to match-to-sample training of baseline relations in the absence of self-reports, the self-report procedure was again implemented during emergent testing. Strong correlations were observed between participants' performance on match-to-sample trials and their verbal reports of that performance. The results observed led to the conclusion that protocol analysis may additionally prove to be of use with regard to the investigation of the qualitative differences between emergent relations suggested by previous research into the effects of nodal distance (e.g., Adams et al., 1993a, 1993b; Fields et al., 1990, 1993). Rehfeldt et al. (1998) have subsequently employed a think-aloud procedure to investigate relations between the formation of equivalence classes composed of elements of compound stimuli and the blocking effect. These researchers, however, reported that participants' overt verbal behaviour during experimentation did not correlate with those participants' non-verbal performance. Confirmations of strong correlations between participants' verbal protocols and their non-verbal behaviour during equivalence research have, however, been reported elsewhere (e.g., Critchfield & Perone, 1990; Potter et al., 1997).

As noted above, there has been considerable variation in the methods employed by behavioural researchers, both in the collection and the analysis of participants' verbal protocols. Most commonly, procedures have been modelled after Ericsson and Simon's (1993) instructions regarding the implementation of think-aloud or "talk-aloud" procedures. Although Ericsson and Simon (1993) have drawn a distinction between the two procedures, the terms have almost always been used interchangeably within the equivalence literature, simply to denote protocol analysis as an experimental technique (Rehfeldt et al., 1998). Because of the differences between behavioural and cognitive theory, the coding of data gathered within equivalence research has not been easily regulated (Critchfield et al., 1998). A feature of coding common across experiments, however, has been the formulation of experimenter-designated response classes to which participants' verbalisations can be assigned. Nevertheless, the nature of these classes has usually been determined both by the experimental preparations employed and the experimenters' theoretical orientation (Austin & Delaney, 1998).

3.5 SUMMARY

Although match-to-sample has proved to be the procedure of choice for equivalence research, a number of variants (simultaneous, delayed, delayed cue, successive, and errorless) have been employed, both in the establishment and testing of equivalence classes. One or more of three main training procedures (many-to-one, one-to-many, and linear) have formed the empirical basis of the majority of equivalence research, complemented by the use of one or more of three main testing procedures (complex-to-simple, simple-to-complex, and simultaneous). Use of procedures other than match-to-sample has also been reported, however: These have included respondent conditioning techniques, simple discrimination training, observational learning, outcome-specific procedures, and sequence training. Methodological considerations have also been highlighted. These include the designations by which stimuli and stimulus classes are referred to, and the accuracy criteria employed during equivalence research. A number of issues regarding the nature and deployment of reinforcers and comparisons have also been noted, as have considerations regarding inter-trial intervals. Various measures of participants' responding have been employed in equivalence research, and considerations relating to the measurement and conceptualisation of reflexivity have also emerged. A number of issues relating to the verbal behaviour of participants and experimenters during experimentation have also been indicated: These have included the nature and effects of experimental instructions, the use of verbal reports as data, and the analysis of participants' concurrent verbal protocols. Chapters Four to Eight present the programme of research that forms the empirical basis of this thesis.

4. STUDY ONE: TESTING THE NAMING HYPOTHESIS

4.1 INTRODUCTION

From its inception, equivalence research has been closely linked with the study of verbal behaviour (Sidman, 1971; Sidman & Cresson, 1973) and the relationship has remained intimate over the years (Hall & Chase, 1991; Hayes & Hayes, 1992; Sidman et al., 1986). During the last decade, however, debate has intensified regarding the primacy of verbal behaviour or equivalence, and over the suggestion that verbal behaviour, and naming in particular, may provide the necessary and sufficient preconditions for equivalence class formation (Horne & Lowe, 1996).

Although the majority of relevant studies have reported no consistent evidence for either the necessity or sufficiency of naming in equivalence class formation (e.g., Green, 1990; Lazar et al., 1984; Sidman et al., 1986), acceptance of such a conclusion has been tempered by the methodological problems inherent in separating verbal and nonverbal experimental outcomes (Perone, 1988; Shimoff, 1984, 1986; Wulfert et al., 1991), especially when the verbal processes implicated may be covert and unmeasurable (Skinner, 1969). Although some studies employing children as participants (Barnes et al., 1990; Devany et al., 1986) have suggested correlations between chronological age, development of verbal behaviour, and equivalence class formation, criticisms have been made on methodological grounds (Saunders & Green, 1996) and because of the correlational nature of the data (Dugdale & Lowe, 1990). As Sidman (1994) has pointed out, causal data would be a step in the right direction toward resolving the naming debate.

With regard to this, Dugdale and Lowe (1990) noted that when normal children who had previously failed equivalence tests were taught a common name for visual stimuli congruent with the classes defining the experiment, they then proceeded to exhibit the pattern of behaviour that defines equivalence. Likewise employing visual stimuli, Eikeseth and Smith (1992) reported the facilitative effects of common naming interventions in remediating failures on tests for equivalence among autistic children. Their findings did not, however, demonstrate the necessity of naming for class formation. Lowe and Beasty (1987) similarly demonstrated the efficacy of teaching intraverbal naming strategies to children in producing equivalence.

Nevertheless, even very young human participants bring to experimentation an extensive verbal history, which in most paradigms has constituted an uncontrolled and uncontrollable variable, precluding definitive analyses. Although various studies have attempted to circumvent such difficulties by using language-disabled participants (e.g., Barnes et al., 1990; Devany et al., 1986) or abstract stimuli (e.g., Lazar et al., 1984; Saunders et al., 1988), other research has capitalised on adult participants' verbal abilities to demonstrate that the pronounceability of textual stimuli can prove indicative of the speed and accuracy with which equivalence classes form (Mandell & Sheen, 1994). Other studies have demonstrated that adult participants' verbal behaviour can either facilitate or hinder class formation, depending on the congruence of the naming strategies employed with the experimenter-designated classes governing positive test outcomes (Bentall et al., 1993; Dickins et al., 1993; Smith et al., 1996). Such studies may not be optimal, however, in that they rely on differences between participants exposed to different stimuli to infer the role of naming, as Remington (1996) has pointed out.

The present research was proposed to elucidate the role of verbal behaviour in equivalence class formation, by presenting different arrangements of the same easily nameable, yet formally unrelated, visual stimuli to verbally able adult participants. As Remington (1996) has observed, if the properties of stimulus names, rather than the stimuli named, determine the ease with which equivalence classes form, it should be possible to identify groupings of stimuli that will be more or less easily related on the basis of their names. Research has suggested that a highly salient feature of words is their phonological characteristics, especially when those characteristics promote rhyme with other words (e.g., Goswami & Bryant, 1990). If participants name visual stimuli, then classes composed of stimuli whose names rhyme might be expected to become equivalent more readily than classes of stimuli whose names share no such similarity.

Three experimental conditions were compared; one in which the names of stimuli forming classes rhymed and two control conditions in which classes were composed of different combinations of the same stimuli whose names were phonologically unrelated. Because the paradigm constituted a typical arbitrary visual-visual paradigm and participants in all conditions were exposed to the same performance-contingent training and testing programme, it was predicted that equivalence would be demonstrated by participants in all conditions. If the hypothesis was correct, however, equivalence would be demonstrated more quickly and with fewer errors by participants in the former condition. As a measure of the normativeness of stimulus names used, written post-experimental tests of naming were conducted. Effects of nodal distance (Fields et al., 1993, 1990; Fields & Verhave, 1987) and the possible generalisation of rhyme-based equivalence classes to novel stimuli whose names rhymed were also assessed.

4.2 METHOD

4.2.1 Participants

Thirty students and staff at the University of Southampton (18 female, 12 male) volunteered to participate in the study and were assigned randomly, but in equal number, to three experimental conditions (rhyme, orthogonal, and diagonal). Aged between 18 and 40 years, all were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 per 30 min, independent of experimental performance. Data from one participant, whose first language was Italian, are excluded from formal analyses and are considered separately.

4.2.2 Apparatus and Setting

Using software designed specifically for equivalence research (Dube & Hiris, 1996), a Power Macintosh® computer presented all stimuli and automatically recorded participants' responses and response latencies. During match-to-sample trials, its 15-in. (38 cm) monitor displayed five transparent "keys" (4.5 cm square), that were indiscernible against a white background. Sample stimuli were presented on the centre key and comparisons on the four outer keys (see Figure 4). During generalisation testing, one of the outer keys, its position varying from trial to trial, always remained blank. Participants were tested individually in a small windowless cubicle (1.5 m by 2.9 m) containing a desk, on which were placed a sheet of written instructions, the computer, monitor, and mouse, and an envelope concealing a pen and post-test booklet for completion subsequent to match-to-sample testing. No keyboard was visible, and responses were made using the mouse. All participants completed the experiment in one sitting, which never exceeded 1.5 hr duration.

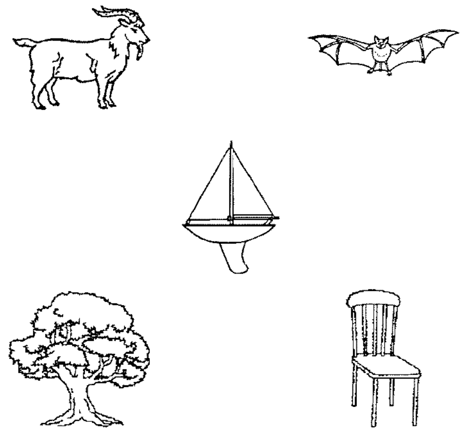

Figure 4. Typical match-to-sample screen display, illustrating an A1B1 trial for the Rhyme condition; *boat* (sample) *goat* (correct comparison).

4.2.3 Stimuli and Class Arrangements

Stimuli were 25 black-and-white pictures of easily nameable items (some adapted from Snodgrass & Vanderwart, 1980), the normative names of which were each between three and five letters in length (see Table 1). Sixteen of these stimuli

Baseline training and Emergent testing				Generalisation testing		
boat	goat	note	coat	can	man	fan
flea	tree	bee	key	dog	frog	log
rat	bat	cat	hat	snake	cake	rake
bear	chair	hair	pear			

Table 1. Normative names of pictorial stimuli used in baseline training, emergent testing, and generalisation testing.

provided the potential for 4 four-member equivalence classes in each condition; the other nine, presented in the final testing phase only, the potential for 3 three-member generalised classes. Although all participants were exposed to the same stimuli throughout the experiment, the arrangements of stimuli composing potential equivalence classes differed between the three conditions. In the rhyme condition, classes were composed of stimuli whose names rhymed with each other, and trials

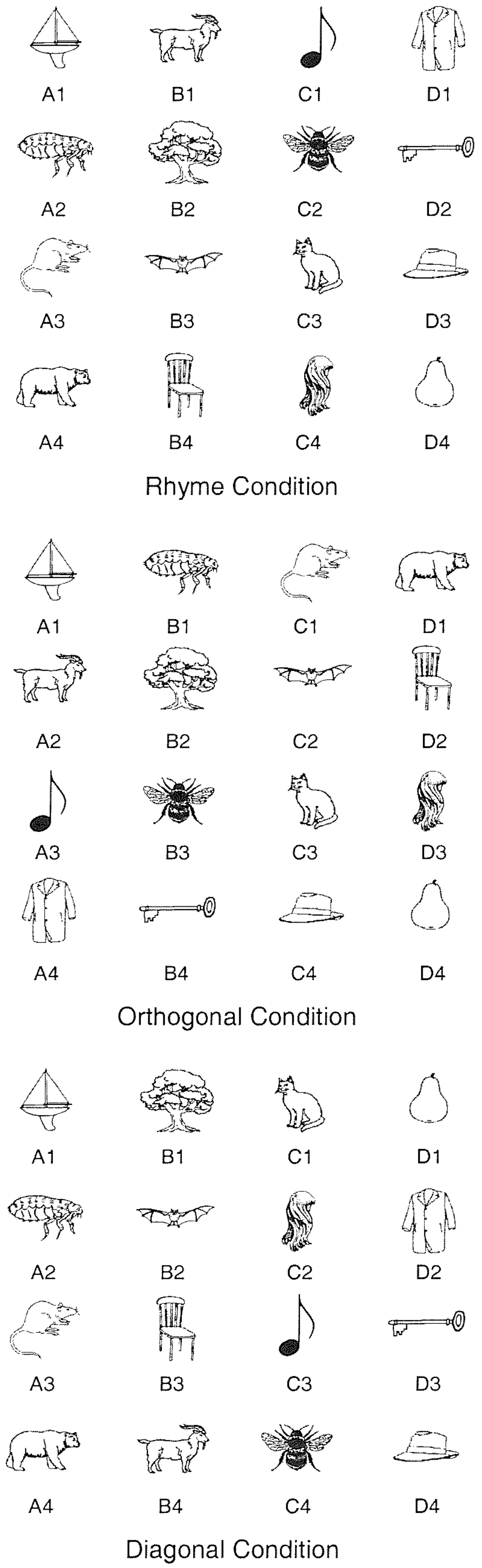

Figure 5. Stimuli and class configurations used in baseline and emergent trials for all conditions. Numbered rows denote classes, lettered columns denote stimuli.

always presented a sample whose name rhymed with that of the correct comparison but never with those of the incorrect comparisons. For two control conditions, classes consisted of stimuli whose names did not rhyme. In the orthogonal condition, class rearrangement was such that all available comparisons' names always rhymed with each other but never rhymed with that of the sample. In the diagonal condition, one of the incorrect comparisons' names always rhymed with the sample's name but never with those of either the correct or the other two incorrect comparisons, whose names also did not rhyme with each other (see Figure 5 for stimuli and class arrangements for all conditions). The other nine stimuli (presented in identical class configurations for all conditions) permitted use of only three comparisons per trial, the name of one of which always rhymed with the sample's name but never with those of the other two comparisons (see Figure 6).

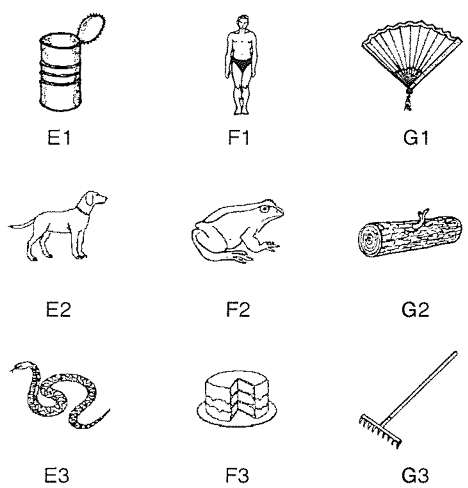

Figure 6. Stimuli and class configurations used in generalisation testing for all conditions. Numbered rows denote classes, lettered columns denote stimuli.

4.2.4 Procedure

Participants were asked to familiarise themselves with the instructions before them, and were then left to complete the experiment.

4.2.4.1 Instructions.

Initially, the following text was displayed on the computer's monitor: "When you are familiar with the written instructions, please click on 'Continue' to start the experiment." The written instructions were as follows:

When the experiment begins, and at the start of each subsequent trial, you will see a picture in the middle of the screen in front of you. Use the mouse to click on it. More pictures will now appear in the corners of the screen. Use the mouse to click on one of these. At first, you will receive feedback on your choices, a "beep" for correct, and a "buzz" for incorrect. During later stages of the experiment, you will no longer receive feedback on your choices--the computer will tell you when. Keep on going however, and continue to do the best you can! Please aim to complete the experiment as quickly and accurately as possible. The computer will record your performance throughout, and a message on screen will tell you when the experiment is over. When you are ready to start, please click on "Continue". Thank you for participating in this experiment. You are free to leave at any point.

The specified action removed the on-screen instructions and match-to-sample training commenced.

4.2.4.2 General procedure and match-to-sample contingencies.

Each trial began with presentation of a sample stimulus, an observing response with the mouse causing comparison stimuli to be displayed. All stimuli remained in view until selection of a comparison caused them immediately to disappear, followed, after a 1-s interval, by presentation of the next trial. Comparison selections made within 0.5 s of presentation had no such consequence, however, and all stimuli remained in view. There was no limit to trial duration. Positions of correct and incorrect comparisons varied pseudo-randomly from trial to trial and, throughout training and testing, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3, B4). At no point did the location of the correct comparison remain constant for more than two consecutive trials, nor did the same sample stimulus appear for more than two trials consecutively.

All participants were exposed to the same performance-contingent training and testing programme. During reinforcement training, selection of class-consistent comparison stimuli was followed by a beep and the word "CORRECT" displayed on the screen. Other choices resulted in a buzz and a darkened screen. During testing without reinforcement, however, the only consequence of a response was the presentation of the next trial. To assess effects of nodal distance, training was sequential (i.e., AB-BC-CD), allowing the potential emergence of 12 symmetric (BA, CB, DC), eight one-node transitive (AC, BD), eight one-node equivalence (CA, DB), four two-node transitive (AD), and four two-node equivalence relations (DA). The overall procedure for all conditions was designed as a series of successive training and testing blocks, the details of which are given below.

4.2.4.3 Phase 1: Establish AB, BC, and CD baseline relations.

Initially, AB relations were trained, with each of the four relations (i.e., A1B1, A2B2, etc.) presented in pseudo-random order once every four trials. When a criterion of 12 consecutive correct responses had been achieved, BC relations were trained in identical fashion. When the same criterion had been attained for these relations, CD relations were established, again to the same criterion.

Trained relations								
AB			BC			CD		
Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
A1	B1	B2, B3, B4	B1	C1	C2, C3, C4	C1	D1	D2, D3, D4
A2	B2	B1, B3, B4	B2	C2	C1, C3, C4	C2	D2	D1, D3, D4
A3	B3	B1, B2, B4	B3	C3	C1, C2, C4	C3	D3	D1, D2, D4
A4	B4	B1, B2, B3	B4	C4	C1, C2, C3	C4	D4	D1, D2, D3

Table 2. Baseline trial configurations, using single-sample and four comparison displays.

4.2.4.4 Phase 2: Review baseline relations, with feedback.

Consequent to fulfilment of the above criteria, all baseline relations were reviewed in 12-trial blocks, with all AB, BC and CD trials intermixed in pseudo-random order. Samples from the same class were never presented consecutively. On completion of one reinforcement trial block with 100% accuracy, the next review phase commenced, assessing baseline maintenance in extinction.

4.2.4.5 Phase 3: Review baseline maintenance, without feedback.

Otherwise identical to Phase 2, all trials during this phase were completed in the absence of reinforcement. If performance remained at 100% accuracy over the first

12-trial block, emergent testing commenced (Phase 4). If criterion was not achieved by the end of the second block, however, baseline relations were reviewed, again with feedback (Phase 2). Review of baseline and its maintenance in extinction continued in this way until 100% accuracy was demonstrated over one block of no-reinforcement trials. Table 2 shows the conditional discriminations established during phases 1 to 3.

4.2.4.6 Phase 4: Emergent testing.

All possible emergent relations except reflexivity were presented, in pseudo-random order, in a maximum of four 36-trial blocks (see Table 3). Generalised

	Emergent relations											
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
Symmetry												
BA	B1	A1	A2, A3, A4	B2	A2	A1, A3, A4	B3	A3	A1, A2, A4	B4	A4	A1, A2, A3
CB	C1	B1	B2, B3, B4	C2	B2	B1, B3, B4	C3	B3	B1, B2, B4	C4	B4	B1, B2, B3
DC	D1	C1	C2, C3, C4	D2	C2	C1, C3, C4	D3	C3	C1, C2, C4	D4	C4	C1, C2, C3
One-node transitivity												
AC	A1	C1	C2, C3, C4	A2	C2	C1, C3, C4	A3	C3	C1, C2, C4	A4	C4	C1, C2, C3
BD	B1	D1	D2, D3, D4	B2	D2	D1, D3, D4	B3	D3	D1, D2, D4	B4	D4	D1, D2, D3
One-node equivalence												
CA	C1	A1	A2, A3, A4	C2	A2	A1, A3, A4	C3	A3	A1, A2, A4	C4	A4	A1, A2, A3
DB	D1	B1	B2, B3, B4	D2	B2	B1, B3, B4	D3	B3	B1, B2, B4	D4	B4	B1, B2, B3
Two-node transitivity												
AD	A1	D1	D2, D3, D4	A2	D2	D1, D3, D4	A3	D3	D1, D2, D4	A4	D4	D1, D2, D3
Two-node equivalence												
DA	D1	A1	A2, A3, A4	D2	A2	A1, A3, A4	D3	A3	A1, A2, A4	D4	A4	A1, A2, A3

Table 3. Emergent trial configurations, using single-sample and four comparison displays.

identity-matching repertoires were assumed (cf. Bush et al., 1989). If participants satisfied the criterion of a minimum of 35 of 36 class-consistent responses in any one emergent testing block, generalisation testing commenced (Phase 5). If, however, criterion had not been achieved at the end of two consecutive blocks of emergent testing, baseline relations were again reviewed, first without feedback (Phase 3) and, if 100% accuracy had not been achieved at the end of two trial blocks (24 trials), again with feedback (Phase 2). Baseline review continued in this way until all relations were again demonstrated with 100% accuracy over one block of test trials. The two final blocks of emergent testing were then presented. Generalisation testing followed completion of the fourth emergent block, regardless of performance.

4.2.4.7 Phase 5: Generalisation testing.

Two consecutive blocks of 18 test trials each were presented, involving previously unseen stimuli. Each trial presented a novel sample stimulus followed, after an observing response, by three novel comparisons (see Table 4). If, at the end of the first block, participants had selected only comparisons whose names rhymed, match-to-sample testing ended. An on-screen message automatically informing participants of this and asked them to complete the post-test in the envelope before them. Following the second block of generalisation trials, the same message was displayed, regardless of performance.

Generalised Emergent Relations									
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
EF	E1	F1	F2, F3	E2	F2	F1, F3	E3	F3	F1, F2
FE	F1	E1	E2, E3	F2	E2	E1, E3	F3	E3	E1, E2
FG	F1	G1	G2, G3	F2	G2	G1, G3	F3	G3	G1, G2
GF	G1	F1	F2, F3	G2	F2	F1, F3	G3	F3	F1, F2
EG	E1	G1	G2, G3	E2	G2	G1, G3	E3	G3	G1, G2
GE	G1	E1	E2, E3	G2	E2	E1, E3	G3	E3	E1, E2

Table 4. Generalisation trial configurations, using single-sample and three comparison displays.

4.2.4.8 Phase 6: Naming post-test.

Subsequent to match-to-sample testing, participants completed a written post-test, that was designed to indicate their naming responses during the experiment. The booklet was headed by the following instructions:

> Printed below are the pictures that you have seen during the experiment. Did you mentally name any of them, or refer to them in any way during testing? If you did, please write under each picture the name, or names, you

used for it during the experiment. If you did not refer to a picture in any such way, please leave the space underneath it blank.

All experimental stimuli were presented, each followed by a blank space and dotted line.

4.3 RESULTS

All participants completed the experiment, although the accuracy with which they did so strongly differentiated participants in the rhyme condition from those in the two control conditions. Individual participants' trials and errors during all phases of match-to-sample training and testing are presented in Appendix A, and their latency data during the first block of emergent testing appear in Appendix B.

4.3.1 Phase 1: Establish Baseline Relations

Acquisition of baseline relations was easiest for participants in the rhyme condition and was most difficult for those in the orthogonal condition. The mean number of trials required by rhyme participants to meet all three criteria for this phase was 53.6 (SD = 24.7) with a mean error score of 9.9 (SD = 12.3), whereas the mean number of trials required by participants in the orthogonal condition was 115.7 (SD = 25.1), with a mean of 42.6 errors (SD = 18.6). Diagonal participants required a mean of 98 trials (SD = 36), with a mean of 31.2 errors (SD = 16.7). With one exception (Participant SG), all participants in the rhyme condition required fewer trials to establish baseline relations than any participant in either of the two control conditions. Participant SG aside, the greatest number of trials required by a participant in the rhyme condition was 57 (nine errors). By contrast, the smallest number of trials required by any participant in the orthogonal condition was 87 (29 errors), and was 67 (21 errors) in the diagonal condition: The smallest number of trials required by a participant in the rhyme condition was 36 (0 errors)--the minimum to meet the criteria. Apart from Participant SG, the maximum number of errors made by anyone in the rhyme condition was 14 and, although the most accurate diagonal participant made only 12 errors, no one else in that condition made less than 20 errors. In the orthogonal condition, the minimum error score was 25. Figure 7 shows the mean number of trials required and errors made by participants in all conditions during this phase of training.

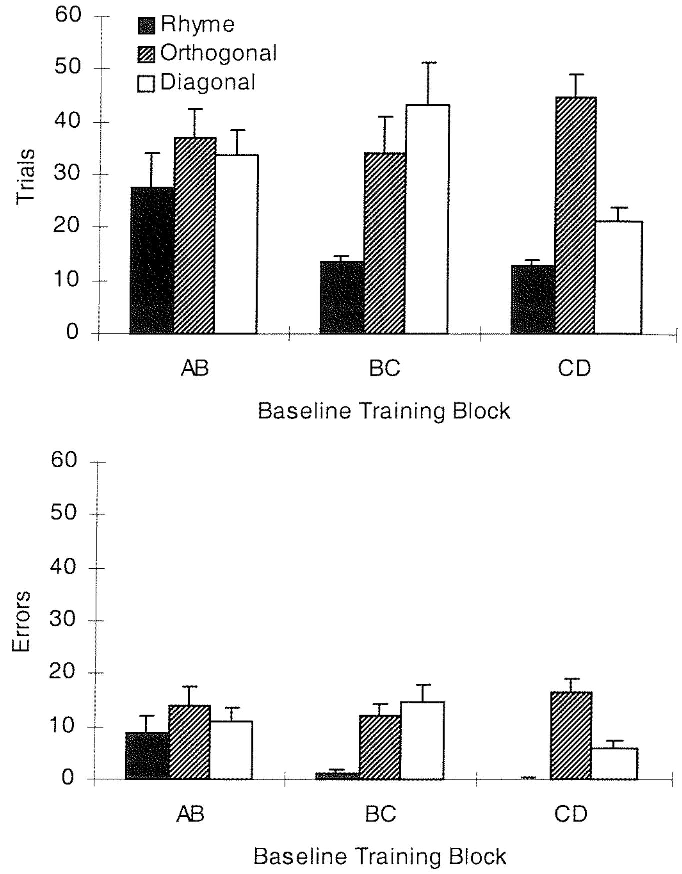

Figure 7. Mean trials and errors (+*SE*) in all conditions to establish component baseline relations (Phase 1).

The trial and error data were subjected to mixed design analyses of variance, in both of which the between-participant factor was condition (rhyme, orthogonal) and the within-participant factor was training block (AB, BC, CD). Regarding the number of trials required to meet the initial criteria, there was a significant main effect of condition, $F(2,27) = 12.11, p < .0001$, and a significant interaction of Condition x Training block, $F(4,54) = 4.96, p < .01$. Regarding errors, there was a significant main effect of condition, $F(2,27) = 10.67, p < .0001$, and a significant interaction of Condition x Training block, $F(4,54) = 3.73, p < .01$. Scheffé tests (at an alpha level of .05) indicated that the number of trials required and errors made during AB training did not differ significantly between conditions. During BC training, however, the number of trials and errors in the rhyme condition was significantly smaller than in the control conditions, which did not differ significantly from each other. Again, during CD training, significantly fewer trials were required in the rhyme condition than in the orthogonal condition, and significantly fewer errors made in the former condition than in the orthogonal and diagonal conditions (the latter at an alpha level of $p = .084$). There were also significantly fewer trials required and errors made in the diagonal condition than in the orthogonal condition.

4.3.2 Phases 2, 3, and 4: Baseline Review and Emergent Testing

4.3.2.1 Review prior to initial emergent testing.

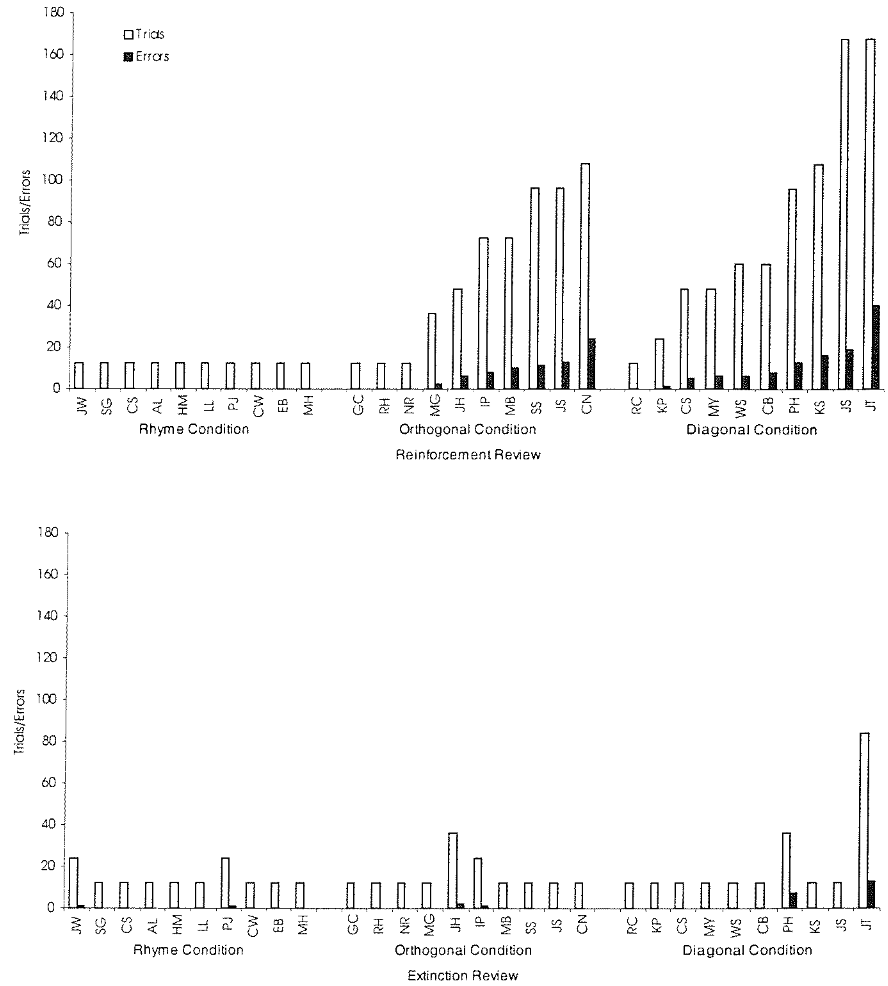

Figure 8. All participants' trials and errors during reinforcement and extinction baseline reviews (phases 2 and 3) prior to initial emergent testing.

All 10 participants in the rhyme condition achieved complete accuracy during their first block of reinforcement review, but only three participants in the orthogonal condition, and one in the diagonal condition, performed likewise. The top half of Figure 8 shows the total number of trials and errors for participants in each condition during this review phase.

Subsequent to meeting criterion for reinforcement review, all participants but two in each condition performed without error during their first block of extinction review and proceeded to emergent testing. Of the remaining participants, both of those in the rhyme condition, and one in the orthogonal condition performed errorlessly during their second block of extinction review. The other orthogonal participant required one additional block of reinforcement review before meeting criterion in extinction. In the diagonal condition, one participant required two additional reinforcement blocks; the other required an additional 27 blocks. [1] The lower half of Figure 8 shows the total number of trials and errors for all participants during extinction review.

4.3.2.2 Initial emergent testing (blocks 1 and 2).

Although all participants showed errorless baseline maintenance in extinction immediately prior to emergent testing, equivalence was confined almost exclusively to participants in the rhyme condition.

Within the first block of emergent testing, all rhyme participants but one fulfilled the 35 of 36 criterion for equivalence; seven performed errorlessly. The remaining participant (SG) performed without error during the second testing block. By contrast, only three participants from the control conditions achieved criterion during initial testing, one from each condition in the first block and another from the diagonal condition in the second block. From the 18th trial of the first block, and throughout the second block, one participant (RC) in the diagonal condition selected only comparisons whose names rhymed with those of the samples. The left sections of Figure 9 show means of percentage error scores made by participants in all conditions during these two blocks. Percentages have been presented because the numbers of emergent relations composing each block were unequal.

These data were subjected to a mixed-design analysis of variance in which the between-participant factor was condition (rhyme, orthogonal, diagonal) and the

---

[1] Of these blocks, three were completed without error. Nine blocks contained a single incorrect response to Sample A3, and another four blocks contained a single incorrect response to Sample B4. A further six blocks contained incorrect responses to both of these samples. None of the remaining five blocks contained more than four errors, and only two of these blocks did not contain an incorrect response to Samples A3, B4, or both.

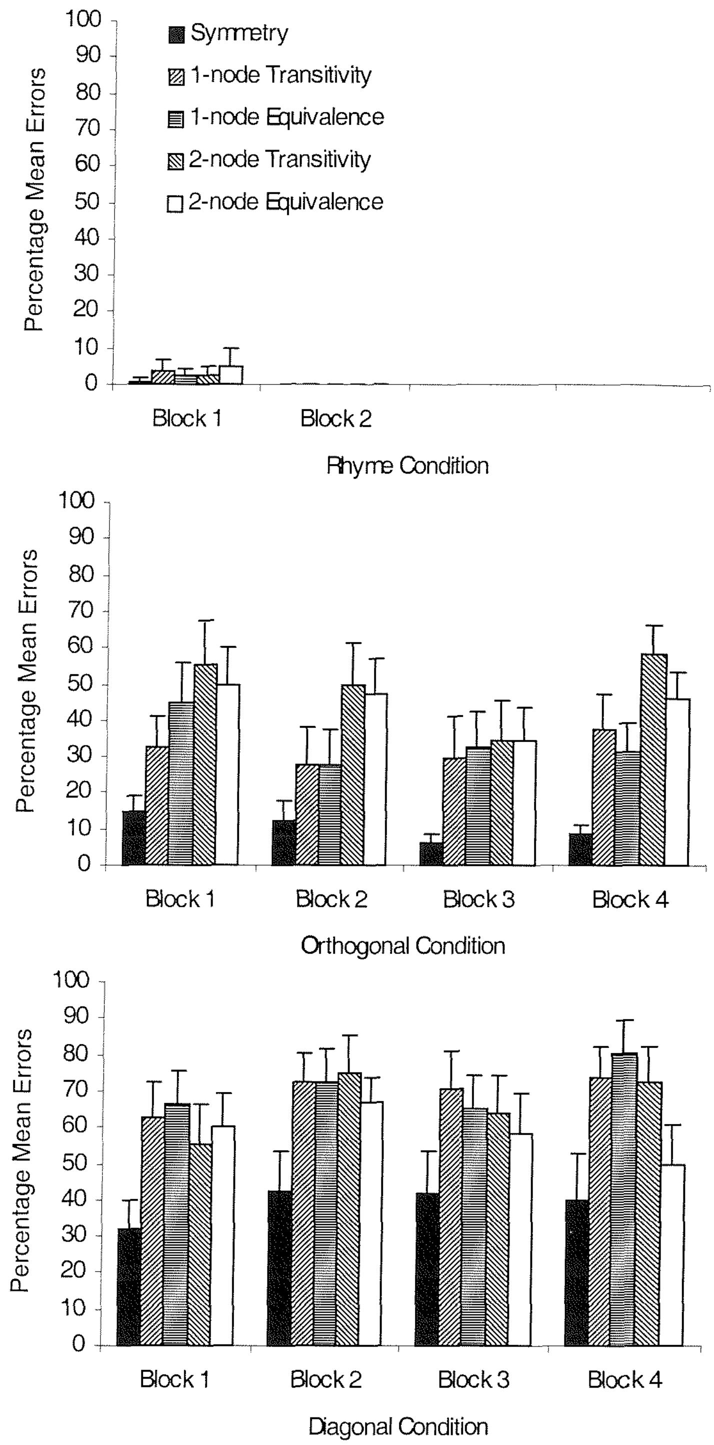

Figure 9. Percentage mean errors (+*SE*) in the Rhyme, Orthogonal, and Diagonal conditions during emergent testing (Phase 4).

within-participant factor was relation type (symmetry, one-node transitivity, one-node equivalence, two-node transitivity, two-node equivalence). Regarding the number of errors made during the first block of emergent testing, there was a significant main effect of condition, $F(2,27) = 14.86, p < .0001$, of relation type

$F(4,108) = 14.41, p < .0001$, and a significant interaction of Condition x Relation Type $F(8,108) = 4.54, p < .0001$. Scheffé tests (at an alpha level of .05) further indicated that although significantly fewer errors were made in the rhyme condition than in either of the control conditions, those conditions did not differ significantly from each other. Post hoc means comparisons (at an alpha level of .05) indicated that participants in both control conditions made significantly fewer errors on symmetry trials than on any other trial type, that participants in the orthogonal condition made significantly fewer errors on trials of one-node transitivity than on any other trial type, and that participants in the diagonal condition made significantly fewer errors on trials of two-node transitivity than on trials of two-node equivalence. No significant differences were observed, however, in the data of the rhyme condition.

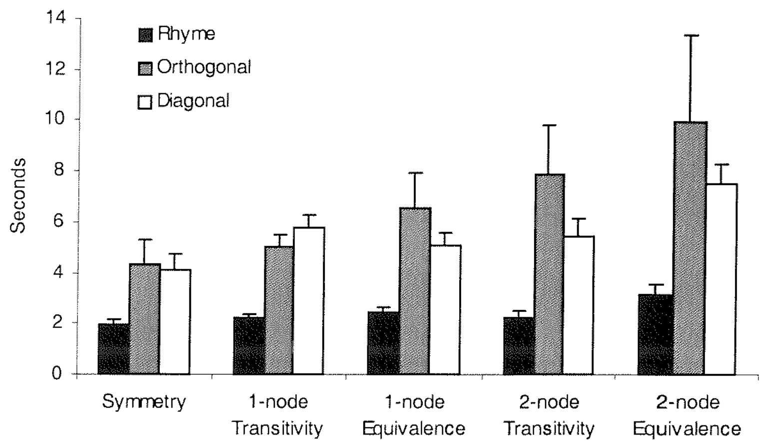

Figure 10. Mean response latencies (+*SE*) in all conditions during the first block of emergent testing (Phase 4).

Response latencies also differentiated the rhyme condition from the control conditions during emergent testing. Mixed design analysis of variance of latencies during the first block of testing--the only block to which all participants were exposed--showed significant main effects of condition, $F(2, 27) = 5.24, p < .05$, and of relation type, $F(4, 108) = 8.48, p < .0001$. An interaction of Condition x Relation Type was also observed, $F(8, 108) = 1.86, p < .07$. [2] Scheffé tests (at an

[2] These differences were confirmed by latency data averaged across all testing blocks to which participants were exposed, analysis of variance again indicating significant main effects of condition, $F(2, 27) = 6.54, p < .01$, and relation type, $F(4, 108) = 15.1, p < .0001$, and a significant interaction of Condition x Relation type, $F(8, 108) = 2.93, p < .01$. Because of the small proportion

alpha level of .05) further indicated that latencies in the rhyme condition were significantly shorter than in the control conditions, and that latencies in the latter conditions did not differ significantly from each other. Post hoc means comparisons additionally indicated that overall, mean latencies were significantly longer on two-node equivalence trials than on any other relation type. Mean latencies for symmetry trials were also significantly shorter than on two-node transitivity trials. Figure 10 shows mean response latencies in each condition during Block 1, for each type of emergent relation tested.

4.3.2.3 Review prior to final emergent testing.

Because all participants in the rhyme condition had shown equivalence during initial testing, none received further baseline review. Of the eight orthogonal and nine diagonal participants who had failed to meet criterion for equivalence, five from each condition demonstrated continued baseline maintenance, performing errorlessly in extinction within two consecutive review blocks. Two of the three remaining orthogonal participants required two additional reinforcement blocks each before meeting criterion in extinction, the other participant a total of six blocks. Two of the four diagonal participants received a single additional review block with reinforcement, the remaining two received a total of four blocks each.

4.3.2.4 Final emergent testing (blocks 3 and 4).

Of the 20 participants in the control conditions, 17 (8 orthogonal and 9 diagonal) received further emergent testing. Of these, only two, from the orthogonal condition, showed equivalence--both in their third testing block. The remaining 15 participants, who never showed equivalence (including Participant RC, who again selected only rhyme comparisons throughout both final blocks), had each received a total of 144 emergent testing trials, and had made an average of 71.8 errors (SD = 9).

The means of percentage error scores of these participants during the final block of emergent testing were subjected to a mixed design analysis of variance in which the between-participant factor was condition and the within-participant factor was trial type. This indicated a significant main effect of condition $F(1,13)$ =

---

of class-consistent responses outside the rhyme condition, analyses were performed on latencies for both correct and incorrect responses throughout.

5.02, $p < .04$, the remaining participants in the orthogonal condition having made fewer errors than those in the diagonal condition. A significant main effect of relation type was also observed $F(4,52) = 7.77$ $p < .0001$, and a significant interaction of Condition x Relation Type $F(4,52) = 2.58$, $p < .05$. Post hoc means comparisons (at an alpha level of .05) indicated that participants in the orthogonal condition made significantly fewer errors on symmetry trials than on any other trial type, and that participants in the diagonal condition made significantly fewer errors on symmetry trials than on any other trial type except two-node equivalence. Participants in the orthogonal condition also made significantly more errors on trials of two-node transitivity than on trials of one-node transitivity or one-node equivalence, whereas participants in the diagonal condition made significantly fewer errors on trials of two-node equivalence than on any other trial type except symmetry. The right sections of Figure 9 show means of percentage error scores made in the orthogonal and diagonal conditions during this phase of testing.

4.3.3 Phase 5: Generalisation Testing

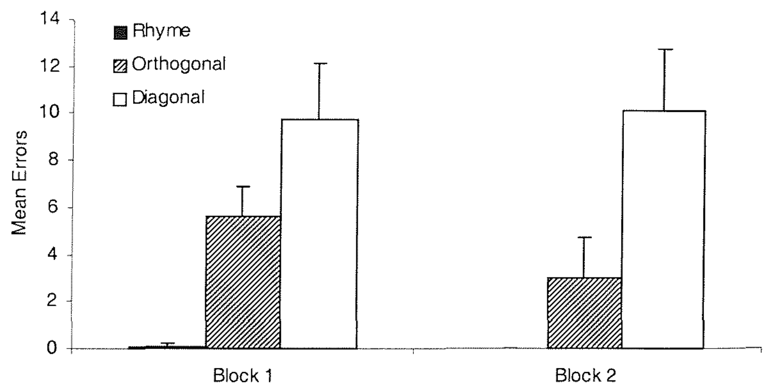

Figure 11. Mean errors (+*SE*) in all conditions during generalisation testing (Phase 5).

Figure 11 shows mean errors made by participants in all conditions during generalisation testing--with errors defined as selection of a comparison stimulus whose name did not rhyme with that of the sample. All but two participants in the rhyme condition achieved complete accuracy in the first block of testing, with the remaining two doing so in their second block (one of these being Participant SG). Only one participant from each of the control conditions made no errors in the first block, although six participants in the orthogonal condition and three in the diagonal condition selected only rhyme comparisons during the second block.

4.3.4 Phase 6: Naming Post-test

All participants' post-tests suggested a high degree of normative stimulus naming, although some stimuli were so named less consistently than others (e.g., *louse*, *bug*, *gnat*, or *mite* for *flea*, *yacht* for *boat*, *mouse* for *rat,* and *tin* for *can*). No rhyme participant indicated using a non-normative stimulus name throughout the experiment, although six indicated that they had changed the names used to the normative ones during the course of the experiment.

One participant in the orthogonal condition named every stimulus normatively, as did three participants in the diagonal condition. Another orthogonal participant reported changing from a non-normative to a normative stimulus name on the basis of rhyme (*oak* to *tree*), and four participants from the control conditions noted using intraverbal phrases (e.g., *there's a bee in your hair*, *the cat eats the pear*, *cat stuck in a tree*, etc.) during training and testing.

4.3.5 Effects of Language Experience

The data from one participant (PM) who was originally assigned to the rhyme condition were rejected from analysis because his first language was Italian. Post-testing revealed his use of a variety of idiosyncratic stimulus names, in both Italian and English (e.g., *sorcio* [*mouse*] for *rat*, *topa* [*beautiful woman*] in response to *hair*, *cap* for *hat*, etc. This was the only participant exposed to rhyme condition training who never met the criterion for equivalence, making a total of 34 errors over the first two emergent blocks and 39 during the latter two. This participant also made a total of 27 errors during generalisation testing, and reported 11 non-normative naming responses on post-testing.

Although her stated first language was English and her data included in the above analyses, it is perhaps worthy of note that Participant SG--whose performance was also exceptional to the rhyme condition--was a fluent speaker of both English and Hindi.

4.4 DISCUSSION

The results suggest that the participants named the experimental stimuli without instruction, and that the phonological properties of the names thus given had influenced match-to-sample performance. When the names of stimuli composing classes rhymed with each other, baseline learning and maintenance, equivalence, and generalised class formation occurred more quickly and reliably than when those names did not rhyme. Less variability was observed in the data of participants who were trained with rhyme combinations of stimuli.

Baseline establishment (Phase 1) consisted of three consecutively presented training blocks (AB, BC, and CD), each composed of newly introduced stimuli, yet all participants but three in the rhyme condition performed without error throughout BC and CD training. Only one participant in the control conditions performed without error during either of these blocks (diagonal Participant CS during CD training). Because the baseline stimuli presented in all conditions bore no consistent formal resemblances to each other, it seems plausible that rhyme participants' consistently class-congruent selection of previously unseen BC and CD comparisons was verbally controlled (Horne & Lowe, 1996) or rule-governed (Skinner, 1969), in that during AB training they had learnt that selection of any comparison whose name rhymed with that of its sample would be correct. Participants in the control conditions, learning a series of purely arbitrary discriminations, had no such straightforward verbal basis for selection available. The virtually errorless maintenance of baseline demonstrated by participants in the rhyme condition during review (phases 2 and 3) did not undermine this interpretation.

It seems equally plausible that the rapid and accurate demonstration of equivalence by participants in the rhyme condition (Phase 4) was a product of the same verbal control: By the end of the second block of 36 trials, all 10 participants in the rhyme condition had met criterion for equivalence, whereas only three of the 20 participants in the control conditions had performed likewise. The data from one participant in the diagonal condition further indicated the functionality of verbal behaviour during the experiment: Participant RC selected only rhyme comparisons throughout his last 137 emergent trials, despite having received no previous reinforcement for selecting such comparisons, and despite having mastered the diagonal condition's baseline. This performance (a flawless demonstration of equivalence had he been in the rhyme condition) indicates that baseline training may be superseded by verbal control during testing without reinforcement, if a ready verbal basis for the categorisation of stimuli is available.

During the first block of emergent testing, no significant effects of nodal distance were observed in the accuracy with which participants in the rhyme condition responded. In accordance with previous research (Kennedy, 1991; Kennedy et al., 1994), however, both control conditions produced significantly greater accuracy on symmetry trials than on those of any other relation type. Also in accordance with previous findings (Fields et al., 1990), participants in the orthogonal condition showed significantly greater accuracy on trials of one-node transitivity than on any other trial type except symmetry. That significantly longer latencies were observed in the responses of all conditions on trials of two-node equivalence than on any other trial type also supported previous reports of the transitivity latency effect (Dickins et al., 1993) in verbally able humans. Although the error scores of the remaining participants in both control conditions differed significantly during the final block of emergent testing, the pattern of errors evident in those scores was similar to those reported for Block 1, suggesting that more extensive exposure to the experimental contingencies would have been required for effects of nodal distance to be minimised (Fields et al., 1993) and for full equivalence to emerge (Lazar et al., 1984; Sidman et al., 1986; Spradlin et al., 1973).

During generalisation testing (Phase 5), all participants in the rhyme condition met the errorless criterion for selection of comparisons whose names rhymed with their samples; all but two did so within the first block of testing. More surprisingly, seven participants in the orthogonal condition and four in the diagonal condition also met this criterion, although only one participant from each condition did so during the first testing block. Not only did the stimuli involved bear no formal resemblance to each other or to the stimuli used in baseline training and emergent testing, but their names were also phonologically unrelated to the names of the baseline stimuli. Additional confirmation that stimuli were named was provided by written post-tests, which also indicated that participants' naming of experimental stimuli was usually normative. Although some participants in the control conditions reported using intraverbal phrases to link stimulus names (as previously reported by Horne & Lowe, 1996), no participants in the rhyme condition reported having used such strategies, although four indicated that they had changed to normative names during the experiment on the basis of rhyme. Anecdotal evidence from participants in the rhyme condition further suggested that they had simply selected "pictures whose names rhymed" early in the experiment.

It is perhaps interesting also to note a correlation between the language experience of two participants exposed to rhyme condition training and testing, and their performance during the experiment. By far the most errors made by any participant in the rhyme condition included in the above analyses were by participant SG who, as noted above, was a fluent speaker of both English and Hindi. Although the post-test suggested her use of mostly normative English names, it does not seem unreasonable to suggest that some of the names she had actually used during the experiment may have been Hindi (cf. Perone, 1988; Shimoff, 1984, 1986). Participant PM, whose first language was Italian, performed at a level akin to the low-accuracy participants in the control conditions. The mixture of normative and non-normative naming responses in both English and Italian that he reported on the post-test strongly suggested that, for him, the task was not one of simply matching stimuli whose names rhymed.

In summary, therefore, the findings of Study One suggest strongly that verbally able humans' performance on equivalence tasks can be influenced by their naming of stimuli and by the phonological properties of the names thus given. Although the present study is clearly not a demonstration of the necessity or even the sufficiency of naming for equivalence or generalised class formation, it nevertheless provides a powerful demonstration that the emergence of untrained relations can be substantially affected both by participants' verbal histories and by their verbal behaviour during experimentation. But in what other ways might that verbal behaviour be functional in the emergence of untrained behaviour?

5. STUDY TWO: NAMING AND CONTEXTUAL CONTROL

5.1 INTRODUCTION

The results of Study One indicated that all participants had named the experimental stimuli without instruction, and that the phonological properties of the names thus given had influenced match-to-sample performance: When the names of stimuli composing classes rhymed with each other, baseline learning and maintenance, equivalence, and generalised class formation occurred more quickly and reliably than when those names did not rhyme. Written post-testing further revealed that although certain stimuli had initially been misnamed by some participants, a high degree of normative stimulus naming had taken place.

Nevertheless, one of the defining characteristics of the experimental analysis of behaviour has been its use of single case designs as an empirical basis for the generation of statements that hold at the level of the individual participant (e.g., Chiesa, 1992, 1994; Sidman, 1960/1988). Would the powerful effects of participants' verbal behaviour observed in Study One also be demonstrable at the individual level? As Remington (1996) has observed, previous research into naming and equivalence (e.g., Mandell & Sheen, 1994) has not been optimal in that it has relied on differences observed between different groups of participants, exposed to different stimuli, to infer the role of naming. Although Study One utilised different combinations of the same stimuli to demonstrate the functionality of naming in the formation and generalisation of equivalence classes, those stimuli were nevertheless presented to different groups of participants. Study Two therefore set out to confirm the findings of Study One using a within-participant design. In so doing, however, an additional empirical question was addressed.

In what manner would participants' naming of stimuli function at the level of the five-term contingency? As noted above, the concept of contextual control has enjoyed a profitable history within the experimental analysis of behaviour, and the formation of contextually controlled equivalence classes has been demonstrated within a variety of experimental preparations (e.g., Barnes et al., 1995; Meehan, 1995; Wulfert et al., 1994; Wulfert & Hayes, 1988), using a variety of experimental stimuli (e.g., Bush et al., 1989; Gatch & Osborne, 1989; Kennedy & Laitinen, 1988; Kohlenberg et al., 1991; Lynch & Green, 1990) presented to a variety of participants (e.g., Lazar & Kotlarchyk, 1986; Meehan & Fields, 1995; Steele & Hayes, 1991).

Although previous research (Car, 1997) has attempted to investigate the relationships between equivalence, naming, and contextual control, the principal aim of this study was to nonetheless to compare the effects of different unprompted and experimenter-induced naming strategies on equivalence class formation, under the contextual control of both positive and negative comparison stimuli. All three experiments of which Carr's (1997) study was composed presented abstract visual stimuli (i.e., Greek letters) to groups of verbally able adults. The difficulties incurred by the use of hard to name visual stimuli have been noted elsewhere (Horne & Lowe, 1996), however, as have the potential confounds introduced by direct naming interventions (e.g., Mandell & Sheen, 1994; Stromer, Mackay, & Remington, 1996). Carr (1997) concluded that although demonstrations of equivalence are subject to contextual control, they are not fundamentally dependent on naming. It might be nonetheless be argued that neither aspect of their study's procedure mentioned above can increase confidence in that conclusion.

No published research has directly investigated the potentially functional role of naming in the formation and generalisation of contextually controlled equivalence classes. The present research was proposed so to do by presenting both rhyme and non-rhyme combinations of the same easily nameable, yet formally unrelated visual stimuli to verbally able adult participants. To facilitate comparison with the findings of Study One, so far as possible, the procedural characteristics of that study were retained. Further to facilitate comparison, the stimuli presented in that study were again employed throughout baseline training and review, emergent testing, and initial generalisation testing: Class arrangements were also identical to those presented to participants in the rhyme and diagonal conditions in Study One. A further experimental question was addressed, however: Would multi-element stimuli, composed of pairs of easily nameable yet novel items whose normative names either did or did not rhyme with each other, prove discriminative for selection of comparisons composed of the contextual stimuli that had previously set the occasion for selection of rhyme or non-rhyme comparisons?

As during Study One, written post-experimental tests of naming were conducted as a measure of the normativeness of participants' stimulus naming. The tests employed were expanded, however, additionally to provide indication of the potential verbal strategies employed by participants during match-to-sample training and testing. As during Study One, effects of nodal distance (Fields et al., 1993, 1990; Fields & Verhave, 1987) were also assessed. To control for possible effects of training order and of the contextual stimuli used, both these factors were counterbalanced.

5.2 METHOD

5.2.1 Participants

Sixteen students and staff at the University of Southampton (nine female, seven male) volunteered to participate in the study. Aged between 18 and 30 years, all were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 per 30 min, independent of experimental performance.

5.2.2 Apparatus and Setting

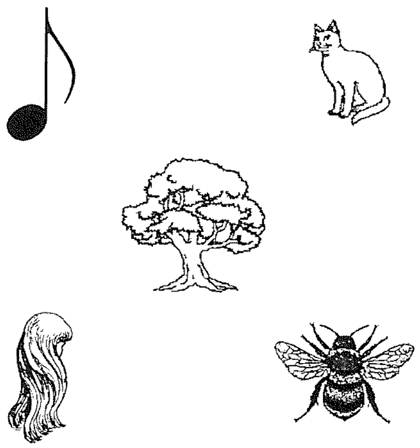

Figure 12. Typical match-to-sample screen display, illustrating a non-rhyme B2C3 trial; *tree* (sample), *cat* (correct comparison).

Using software specifically designed for equivalence research (Dube & Hiris, 1996), a Power Macintosh® computer presented all stimuli and automatically recorded participants' responses and response latencies. Throughout baseline training, and emergent and generalisation testing, its 15-in. (38-cm) monitor displayed five transparent keys (4.5 cm square) that were indiscernible against a red or blue background. Sample stimuli were presented on the centre key, and comparisons appeared on the four outer keys (see Figure 12). During initial generalisation testing, one of the outer keys, its position varying from trial to trial, always remained blank. During secondary generalisation testing, the five transparent keys were presented against a white background and two of the outer keys, their positions varying from trial to trial, always remained blank. Participants were tested individually in a small windowless cubicle (1.5 m by 2.9 m) containing a desk, on which were placed a sheet of written instructions, the computer, monitor, and mouse, and an envelope concealing a pen and post-test booklet for completion subsequent to match-to-sample testing. No keyboard was visible, and responses were made using the mouse. All participants completed the experiment in one sitting, which never exceeded 1.5 hr duration.

5.2.3 Stimuli and Class Arrangements

Baseline training and emergent testing				Initial generalisation testing			Secondary generalisation testing	
boat	goat	note	coat	can	man	fan	well	bell
flea	tree	bee	key	dog	frog	log	sock	lock
rat	bat	cat	hat	snake	cake	rake	spoon	moon
bear	chair	hair	pear					

Table 5. Normative names of pictorial stimuli used in baseline training, emergent and initial generalisation testing, and elements of stimulus compounds used in secondary generalisation testing.

Experimental stimuli were 31 black line drawings of easily nameable items (some adapted from Snodgrass & Vanderwart, 1980), the normative names of which were each between three and five letters in length (see Table 5), and two coloured keys (each of the same colour as one of the contextual stimuli). Contextual stimuli were provided by the red or blue background colours upon which the experimental stimuli were presented. Sixteen of the pictorial stimuli provided the potential for eight contextually controlled four-member equivalence classes, and nine the potential for six contextually controlled generalised classes. The other six stimuli and the two coloured keys, presented during the final testing phase only, provided the potential for two further three-member generalised classes. The former 25 pictorial stimuli had been presented previously in Study One.

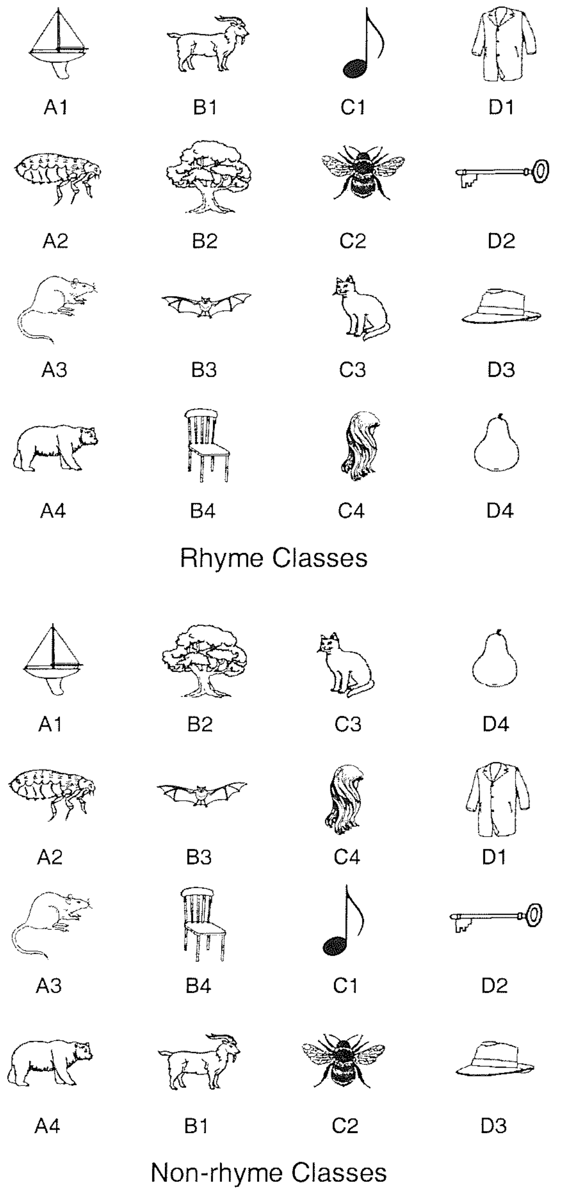

Figure 13. Stimuli and class configurations used in rhyme and non-rhyme baseline and emergent trials. Rows denote classes; columns denote stimuli.

Four stimulus classes (rhyme classes) were trained in the presence of one of the contextual stimuli (rhyme contextual stimulus), and were composed of stimuli whose names rhymed with each other. Trials involving these classes (rhyme trials) always presented a sample whose name rhymed with that of the correct comparison, but never with those of the incorrect comparisons. The other four classes (non-rhyme classes) were trained in the presence of the other contextual stimulus (non-rhyme contextual stimulus), and were composed of stimuli whose names did not rhyme with each other. Trials involving these classes (non-rhyme trials) always presented a sample whose name rhymed with that of one of the incorrect comparisons, but never with those of either the correct or the other two incorrect comparisons, whose names also did not rhyme with each other (see Figure 13 for stimuli and class arrangements). The nine stimuli presented during

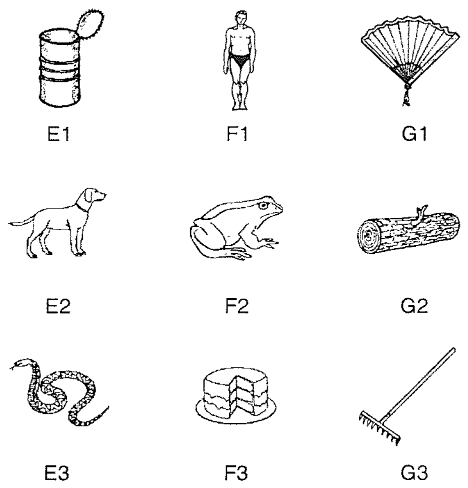

Figure 14. Stimuli used in initial generalisation testing. Numbered rows denote rhyme classes; lettered columns denote stimuli.

initial generalisation testing permitted use of only three comparisons per trial, the name of one of which always rhymed with the sample's name but never with those of the other two comparisons (see Figure 14). The six pictorial stimuli presented during secondary generalisation testing were arranged into two classes of three stimulus compounds (see Figure 15), one of which was composed of pairs of stimuli

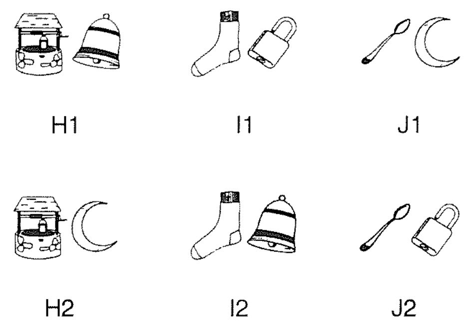

Figure 15. Stimulus compounds used in secondary generalisation testing. Numbered rows denote classes; lettered columns denote stimulus compounds.

whose names rhymed with each other (rhyme compounds), and the other of which was composed of pairs of stimuli whose names did not rhyme with each other (non-rhyme compounds). Each trial presented a stimulus compound as sample, and both coloured keys as comparisons.

Although all participants were exposed to the same experimental and contextual stimuli throughout the experiment, the order of baseline training (rhyme or non-rhyme classes first) and the colour of the rhyme and non-rhyme contextual stimuli (red or blue) were counterbalanced: Four participants were initially trained with rhyme classes under the control of a red background, and four participants with those classes under the control of a blue background. A further four participants were initially trained with non-rhyme classes under the contextual control of a red background, and a final four participants with non-rhyme classes under the contextual control of a blue background. All participants were assigned on a random basis.

5.2.4 Procedure

Participants were asked to familiarise themselves with the instructions before them, and were then left to complete the experiment.

5.2.4.1 Instructions.

Initially, the following text was displayed on the computer's monitor: "When you are familiar with the written instructions, please click on 'Continue' to start the experiment." The written instructions were as follows:

> When the experiment begins, and at the start of each subsequent trial, you will see a picture in the middle of the screen in front of you. Use the mouse to click on it. More pictures will now appear in the corners of the screen. Use the mouse to click on one of these. At first, you will receive feedback on your choices, a "beep" for correct, and a "buzz" for incorrect. During later stages of the experiment, you will no longer receive feedback on your choices--the computer will tell you when. Keep on going however, and continue to do the best you can! Please aim to complete the experiment as quickly and accurately as possible. The computer will record your performance throughout, and a message on screen will tell you when the experiment is over. When you are ready to start, please click on "Continue". Thank you for participating in this experiment. You are free to leave at any point.

The specified action removed the on-screen instructions and match-to-sample training commenced.

5.2.4.2 General procedure and match-to-sample contingencies.

Each trial began with presentation of a sample stimulus, an observing response with the mouse causing comparison stimuli additionally to be displayed. All stimuli remained in view until selection of a comparison caused them immediately to disappear, followed, after a 1-s interval, by presentation of the next trial. Comparison selections made within 0.5 s of presentation had no such consequence, however, and all stimuli remained in view. There was no limit to trial duration. Positions of correct and incorrect comparisons varied pseudo-randomly from trial to trial and, throughout baseline training and emergent and generalisation testing, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3, B4). At no point did the location of the correct comparison remain constant for more than two consecutive trials, nor did the same sample stimulus appear for more than two trials consecutively.

All participants were exposed to the same performance-contingent training and testing programme. During reinforcement training, selection of class-consistent comparison stimuli was followed by a beep and the word "CORRECT" displayed on the screen. Other choices resulted in a buzz and a darkened screen. During testing without reinforcement, however, the only consequence of a response was the presentation of the next trial. To assess effects of nodal distance, training was sequential (i.e., AB-BC-CD), allowing the potential emergence of 24 symmetric (BA, CB, DC), 16 one-node transitive (AC, BD), 16 one-node equivalence (CA, DB), eight two-node transitive (AD), and eight two-node equivalence relations (DA). The overall procedure was designed as a series of successive training and testing blocks: The details presented below are for those participants trained with rhyme classes first. For participants trained with non-rhyme classes first, phases 4 to 6 were presented prior to phases 1 to 3.

5.2.4.3 Phase 1: Establish rhyme baseline relations.

Initially, rhyme AB relations were trained, with each of the four relations (i.e., A1B1, A2B2, etc.) presented in pseudo-random order once every four trials. When a criterion of 12 consecutive correct responses had been achieved, rhyme BC relations were trained in identical fashion. When the same criterion had been attained for these relations, rhyme CD relations were established, again to the same criterion.

5.2.4.4 Phase 2: Review rhyme baseline relations, with feedback.

Consequent to fulfilment of the above criteria, the baseline relations established during Phase 1 were reviewed in 12-trial blocks, with AB, BC, and CD rhyme trials intermixed in pseudo-random order. Samples from the same class were never presented consecutively. On completion of one reinforcement trial block with 100% accuracy, the next review phase commenced, assessing maintenance of rhyme baseline relations in extinction.

5.2.4.5 Phase 3: Review rhyme baseline relations, without feedback.

Trained relations (rhyme)								
AB			BC			CD		
Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
A1	B1	B2, B3, B4	B1	C1	C2, C3, C4	C1	D1	D2, D3, D4
A2	B2	B1, B3, B4	B2	C2	C1, C3, C4	C2	D2	D1, D3, D4
A3	B3	B1, B2, B4	B3	C3	C1, C2, C4	C3	D3	D1, D2, D4
A4	B4	B1, B2, B3	B4	C4	C1, C2, C3	C4	D4	D1, D2, D3

Trained relations (non-rhyme)								
AB			BC			CD		
Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
A1	B2	B1, B3, B4	B1	C2	C1, C3, C4	C1	D2	D1, D3, D4
A2	B3	B1, B2, B4	B2	C3	C1, C2, C4	C2	D3	D1, D2, D4
A3	B4	B1, B2, B3	B3	C4	C1, C2, C3	C3	D4	D1, D2, D3
A4	B1	B2, B3, B4	B4	C1	C2, C3, C4	C4	D1	D2, D3, D4

Table 6. Baseline trial configurations, using single-sample and four comparison displays.

Otherwise identical to Phase 2, all trials during this phase were completed in the absence of explicit reinforcement (test trials). If performance remained at 100% accuracy over the first 12-trial block, training of non-rhyme baseline relations commenced (Phase 4). If criterion was not achieved by the end of the second block, however, rhyme baseline relations were reviewed, again with feedback (Phase 2). Review of rhyme baseline relations and their maintenance in extinction continued in this way until 100% accuracy was demonstrated over one block of test trials. The top section of Table 6 shows the conditional discriminations established during phases 1 to 3.

5.2.4.6 Phases 4, 5, and 6: Establish and review non-rhyme baseline relations.

Phase 4 trained non-rhyme baseline relations in exactly the same way that rhyme baseline relations were established during Phase 1. Likewise, phases 5 and 6 reviewed non-rhyme relations in the same way that rhyme baseline relations were reviewed during phases 2 and 3. Subsequent to fulfilment of these criteria, the next review phase commenced, assessing maintenance of all rhyme and non-rhyme baseline relations in extinction. The lower section of Table 6 shows the conditional discriminations established during phases 4 to 6.

5.2.4.7 Phase 7: Full review of baseline relations.

Initially, all baseline relations established during phases 1 to 6 were reviewed in extinction. All rhyme and non-rhyme AB, BC, CD trials were presented intermixed in pseudo-random order in 24-trial blocks. No more than two trials involving rhyme or non-rhyme baseline relations were ever presented consecutively. If participants satisfied the criterion of a minimum of 22 of 24 correct responses, emergent testing commenced (Phase 8). If criterion was not met in this block, however, all baseline relations were again reviewed in the same manner, but with feedback. On completion of one block of reinforcement trials to the same criterion, maintenance of all baseline relations was again assessed in extinction. Review of baseline and its maintenance continued in this way until criterion was met in extinction.

5.2.4.8 Phase 8: Emergent testing.

All possible emergent relations except reflexivity were presented, in pseudo-random order, in a maximum of two 72-trial blocks (see Table 7). Generalised identity-matching repertoires were assumed. Rhyme and non-rhyme trials were never presented more than twice consecutively. If participants satisfied the criterion of a minimum of 70 of 72 class-consistent responses in the first emergent testing block, initial generalisation testing commenced (Phase 9). If, however, criterion had not been achieved in that block, all rhyme and non-rhyme baseline relations were again reviewed intermixed (Phase 7), first without feedback and, if criterion had not been achieved at the end of one 24-trial block, again with feedback. Review of baseline continued in this way until all relations were again demonstrated to criterion over one block of test trials. The final block of emergent testing was then presented followed, regardless of participants' performance, by initial generalisation testing.

Emergent relations (rhyme)												
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
Symmetry												
BA	B1	A1	A2, A3, A4	B2	A2	A1, A3, A4	B3	A3	A1, A2, A4	B4	A4	A1, A2, A3
CB	C1	B1	B2, B3, B4	C2	B2	B1, B3, B4	C3	B3	B1, B2, B4	C4	B4	B1, B2, B3
DC	D1	C1	C2, C3, C4	D2	C2	C1, C3, C4	D3	C3	C1, C2, C4	D4	C4	C1, C2, C3
One-node transitivity												
AC	A1	C1	C2, C3, C4	A2	C2	C1, C3, C4	A3	C3	C1, C2, C4	A4	C4	C1, C2, C3
BD	B1	D1	D2, D3, D4	B2	D2	D1, D3, D4	B3	D3	D1, D2, D4	B4	D4	D1, D2, D3
One-node equivalence												
CA	C1	A1	A2, A3, A4	C2	A2	A1, A3, A4	C3	A3	A1, A2, A4	C4	A4	A1, A2, A3
DB	D1	B1	B2, B3, B4	D2	B2	B1, B3, B4	D3	B3	B1, B2, B4	D4	B4	B1, B2, B3
Two-node transitivity												
AD	A1	D1	D2, D3, D4	A2	D2	D1, D3, D4	A3	D3	D1, D2, D4	A4	D4	D1, D2, D3
Two-node equivalence												
DA	D1	A1	A2, A3, A4	D2	A2	A1, A3, A4	D3	A3	A1, A2, A4	D4	A4	A1, A2, A3

Table 7. Emergent trial configurations, using single-sample and four comparison displays.

5.2.4.9 Phase 9: Initial generalisation testing.

One block of 36 test trials was presented, involving previously unseen stimuli. Each trial presented a novel sample stimulus followed, after an observing response, by three novel comparisons (see Table 8). Eighteen trials presented stimuli in the presence of the rhyme contextual stimulus, and 18 trials presented the same stimuli in the presence of the non-rhyme contextual stimulus. All trials were intermixed in pseudo-random order and the same contextual stimulus was never presented on more than two consecutive trials. Regardless of participants' performance, secondary generalisation testing commenced subsequent to completion of this block.

Generalised emergent relations (rhyme)									
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
EF	E1	F1	F2, F3	E2	F2	F1, F3	E3	F3	F1, F2
FE	F1	E1	E2, E3	F2	E2	E1, E3	F3	E3	E1, E2
FG	F1	G1	G2, G3	F2	G2	G1, G3	F3	G3	G1, G2
GF	G1	F1	F2, F3	G2	F2	F1, F3	G3	F3	F1, F2
EG	E1	G1	G2, G3	E2	G2	G1, G3	E3	G3	G1, G2
GE	G1	E1	E2, E3	G2	E2	E1, E3	G3	E3	E1, E2

Generalised emergent relations (non-rhyme)									
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
EF	E1	F2 or F3	F1	E2	F1 or F3	F2	E3	F1 or F2	F3
FE	F1	E2 or E3	E1	F2	E1 or E3	E2	F3	E1 or E2	E3
FG	F1	G2 or G3	G1	F2	G1 or G3	G2	F3	G1 or G2	G3
GF	G1	F2 or F3	F1	G2	F1 or F3	F2	G3	F1 or F2	F3
EG	E1	G2 or G3	G1	E2	G1 or G3	G2	E3	G1 or G2	G3
GE	G1	E2 or E3	E1	G2	E1 or E3	E2	G3	E1 or E2	E3

Table 8. Initial generalisation trial configurations, using single-sample and three comparison displays.

5.2.4.10 Phase 10: Secondary generalisation testing.

A maximum of two consecutive blocks of 12 test trials each were presented, involving previously unseen stimuli. Each trial presented a novel stimulus compound as sample followed, after an observing response, by both coloured keys as comparisons (see Table 9). If, at the end of the first block, participants had selected only rhyme keys (i.e., keys of the same colour as the rhyme contextual stimulus) in response to rhyme compounds, and non-rhyme (i.e., keys of the same colour as the non-rhyme contextual stimulus) in response to non-rhyme compounds, match-to-sample testing ended. An on-screen message automatically informed participants of this, and asked them to complete the post-test in the envelope before them. Following the second block of secondary generalisation testing, the same message was displayed, regardless of performance.

Generalised emergent relations		
Sa	Co+	Co-
H1	rhyme key	non-rhyme key
I1	rhyme key	non-rhyme key
J1	rhyme key	non-rhyme key
H2	non-rhyme key	rhyme key
I2	non-rhyme key	rhyme key
J2	non-rhyme key	rhyme key

Table 9. Secondary generalisation testing trial configurations, using single-sample two comparison displays.

5.2.4.11 Phase 11: Naming post-test.

Subsequent to match-to-sample testing, participants completed a written post-test that was designed to elucidate their verbal behaviour during the experiment. The booklet was headed by the following instructions:

> Printed below are the pictures that you have seen during the experiment. Did you mentally name any of them, or refer to them in any way during testing? If you did, please write under each picture the name, or names, you used for it during the experiment. If you did not refer to a picture in any such way, please leave the space underneath it blank.

Initially, the experimental stimuli used during baseline and emergent testing were presented, each followed by a blank space and dotted line. Beneath this was a blank space headed by the following instructions:

> Please use the space below to describe briefly any mental strategies, rules, or other "tricks" that you may have used to learn and remember which of the above pictures went together. If the strategies you used differed depending on whether the screen was RED or BLUE, please describe how

they differed. If you did not use any such strategies, please leave the space blank and complete the rest of the post-test.

Following this, the experimental stimuli used during initial generalisation testing were presented, each followed by a blank space and dotted line. Above this were the following instructions:

> Later in the experiment, you saw the following pictures. As before, please note beneath each any name, or names, by which you may have referred to it during testing.

Next followed a blank space headed by the following instructions:

> In this later stage of the experiment, did you use any mental strategies, "tricks", or rules in choosing which pictures went together? If you did, please enter a brief description below, and as before indicate whether and how they may have differed when the screen was RED and when it was BLUE. If you used no such strategies, please leave the space below blank and complete the rest of the post-test.

Lastly, the stimuli six elements of the stimulus compounds used during secondary generalisation testing were presented separately, each followed by a blank space and dotted line. Above these were the following instructions:

> In the final stage of the experiment, you saw the following pictures. Once again, please note beneath each any name, or names, by which you may have referred to it during testing.

Beneath this was a blank space, headed by the following instructions:

> In this final stage of testing, you chose either a RED or a BLUE square on every trial. Did you use any strategies, "tricks", or rules to make your choices? If you did, please enter a brief description in the space on the next page. If not, leave the space blank and complete the rest of the post-test.

5.3 RESULTS

Although all participants completed the experiment, none demonstrated the emergent relations indicative of the formation of contextually controlled equivalence classes. Individual participants' trial and error scores during all phases of match-to-sample training and testing are presented in Appendix C, and their latency data during emergent testing appear in Appendix D.

5.3.1 Phases 1 and 4: Establish Rhyme and Non-rhyme Baseline Relations

Acquisition of baseline relations was easier when the names of stimuli composing baseline relations rhymed with each other. The mean number of trials required by all participants to meet the three criteria for rhyme baseline relations during this phase of training was 15 (*SD* = 5.2), with a mean of 1.9 errors (*SD* = 3.5). The mean number of trials required by participants to meet the same criteria for non-rhyme relations was 32.6 (*SD* = 18.5), however, with a mean of 10.3 errors (*SD* = 9.5). All participants but one (RR) made fewer errors and required fewer trials to establish rhyme baseline relations than to establish non-rhyme baseline relations. Participant RR aside, the greatest number of trials required by any participant to establish rhyme baseline relations was 54 (13 errors). By contrast, the smallest number of trials required by any participant to establish non-rhyme baseline relations was 54 (13 errors) and the greatest was 170 (79 errors). The smallest number of trials required to establish rhyme baseline relations (participants FT, HL, and JW2) was 36 (0 errors)--the minimum to meet the criteria. Figure 16 shows the mean number of trials required and errors made by all participants during this phase of training.

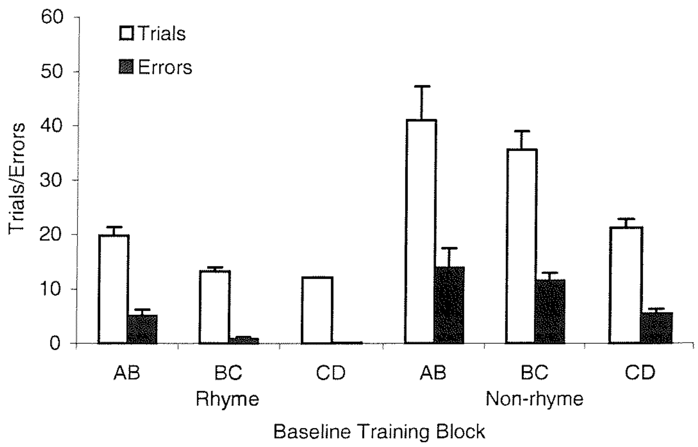

Figure 16. Mean trials and errors (+*SE*) to establish rhyme and non-rhyme baseline relations (phases 1 and 4).

The trial and error data were subjected to two within-participant analyses of variance, in both of which the within-participant factors were phonology (rhyme, non-rhyme) and training block (AB, BC, CD). Regarding the number of trials required to meet the initial criteria, there was a significant main effect of phonology, $F(1, 15) = 40.2$, $p < .0001$, and a significant main effect of training block, $F(2, 30) = 11.3$, $p < .0001$. Regarding errors, there was also a significant main effect of phonology, $F(1, 15) = 41.9$, $p < .0001$, and a significant main effect of training block, $F(2, 30) = 7.6$, $p < .01$. Participants made significantly fewer errors and required significantly fewer trials to establish rhyme baseline relations. Post hoc means comparisons (at an alpha level of .05) indicated that participants both required significantly fewer trials and made significantly fewer errors during the CD training blocks than in either the AB or BC training blocks, and that neither the number of trials required nor the number of errors made in the former blocks differed significantly from each other.

5.3.2 Phases 2, 3, 5, and 6: Rhyme and Non-rhyme Baseline Reviews

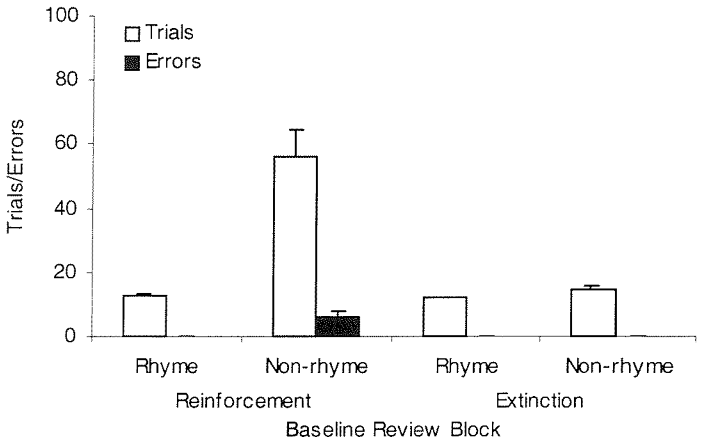

Figure 17. Mean trials and errors (+*SE*) during reinforcement and extinction rhyme and non-rhyme baseline reviews (phases 2, 3, 5, and 6).

All participants achieved complete accuracy during the first reinforcement review block of rhyme baseline relations except participant RW, who made one error. This participant performed without error during the second block. By contrast, only one participant (JW2) performed with complete accuracy during the first reinforcement review block of non-rhyme baseline relations and only three others (RW, AW, and CB) did so during the second. The greatest number of trials required by any participant (JP) during review of non-rhyme relations was 120 (17 errors). The left sections of Figure 17 show the mean number of trials required and errors made by all participants during reinforcement rhyme and non-rhyme baseline reviews.
All participants achieved complete accuracy during the first extinction review block of rhyme baseline relations, and all participants but three (JW1, FT, and AW)

selected only correct comparisons during the first extinction review block of non-rhyme baseline relations. These participants each made one error during the first review block, but performed with complete accuracy during the second block. The right sections of Figure 17 show the mean number of trials required and errors made by all participants during extinction rhyme and non-rhyme baseline reviews.

5.3.3 Phases 7 and 8: Full Baseline Reviews and Emergent Testing

5.3.3.1 Review prior to initial emergent testing.

Despite previously satisfying the criteria for maintenance of rhyme and non-rhyme baseline relations in extinction, only four participants (JP, RW, AW, and AA2) met the 22 of 24 criterion during the first block of extinction full baseline review, although two others (AA1 and AJ) made only three errors each. Of the remaining 10 participants, two (RR and CB) made a total of 11 errors each, seven others (JW1, FT, MS, JW2, AT, AH, and AC) a total of 12 errors, and one (HL) a total of 13 errors. Of the eight participants who had demonstrated maintenance of rhyme baseline relations immediately prior to full baseline review, none made any errors on rhyme trials except JP, who made one. Of the errors made by the four of these participants who did not meet criterion during the first block of extinction baseline review, all but two were other-context selections (i.e., selection of a comparison that would have been correct in the presence of the other-contextual stimulus). Of the eight participants who had demonstrated maintenance of non-rhyme baseline relations immediately prior to full baseline review, none made any errors on non-rhyme trials except HL, who made one. Of the errors made by these participants on rhyme trials, all but two were other-context selections. None of the 12 participants who failed to meet criterion during the first extinction block required more than two additional reinforcement full baseline review blocks to re-establish baseline relations to criterion except participant AC, who required three blocks. Participants made, on average, 7.7 ($SD = 4.3$) and 1.6 ($SD = 2.8$) errors during the first and second reinforcement blocks respectively. All 12 participants satisfied the criterion for full baseline review during the second extinction block and proceeded to emergent testing (Phase 9). The left sections of Figure 18 show total errors made by all participants during both blocks of extinction full baseline review prior to initial emergent testing.

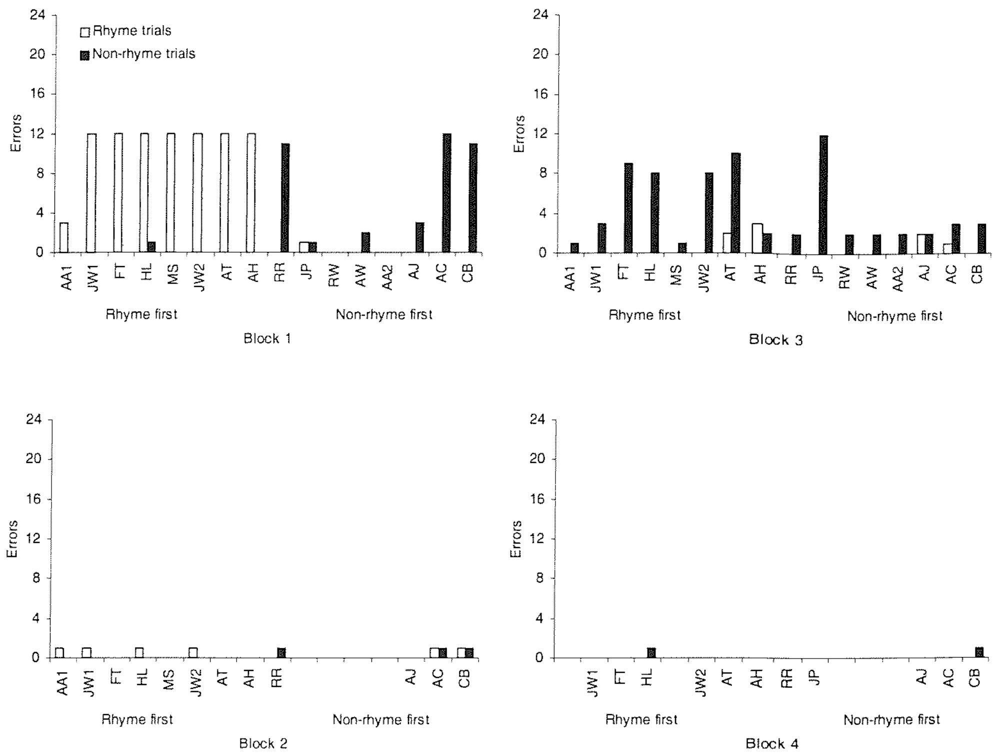

Figure 18. All participants' errors on rhyme and non-rhyme trials during extinction full baseline reviews (Phase 7). Absence of initials indicates that trials were not presented to that participant.

5.3.3.2 Initial emergent testing (block 1).

Although all participants demonstrated full baseline maintenance to criterion in extinction immediately prior to initial emergent testing, none fulfilled the 70 of 72 criterion for contextually controlled equivalence. During the first block of emergent testing, participants made on average 30.8 errors (*SD* = 5.1), but made on average only 1.7 errors (*SD* = 1.6) on the 36 rhyme trials. Only two participants (AH and AJ) made more than three errors on rhyme trials, and 10 participants made one error or less. On average, 1.1 (*SD* = 1.5) of these errors were other-context selections. On the 36 trials of non-rhyme emergent relations, however, participants made on average 29.1 errors (*SD* = 5.7). Only two participants (RR and AJ) made fewer than 25 errors, and 10 participants made 31 or more. On average, 23.8 (*SD* = 10.4) of these errors were other-context selections. The top half of Figure 19 shows each participant's errors on non-rhyme trials, and the number of other-context selections on those trials.

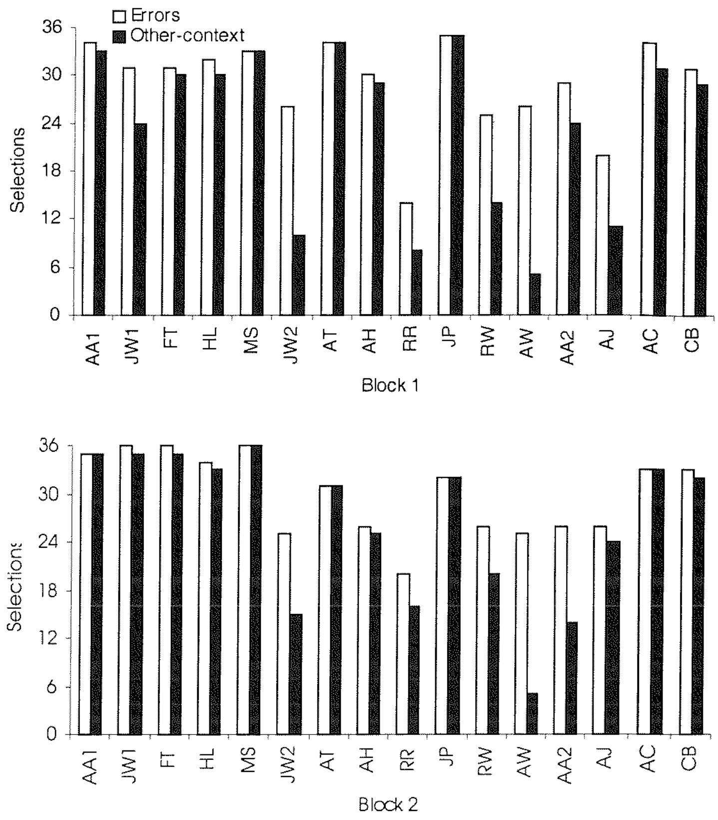

Figure 19. All participants' errors and other-context selections on non-rhyme trials during emergent testing (Phase 8).

Participants' percentage error scores were subjected to a within-participant analysis of variance in which the factors were phonology (rhyme, non-rhyme) and relation type (symmetry, one-node transitivity, one-node equivalence, two-node transitivity, two-node equivalence). There was a significant main effect of phonology, $F(1, 15) = 427.87$, $p < .0001$, and of relation type, $F(4, 60) = 13.29$, $p < .0001$, and a significant interaction of Phonology x Relation Type, $F(4, 60) = 18.03$, $p < .0001$. Participants made significantly fewer errors on rhyme trials. Although post hoc means comparisons (at an alpha level of .05) further indicated that participants made significantly fewer errors on non-rhyme symmetry trials than on non-rhyme trials of any other relation type, no significant differences were observed in participants' error scores on rhyme trials. The top half of Figure 20 shows means of percentage error scores for all participants during the first block of emergent testing. Percentages have been presented because the numbers of emergent relations composing each block were unequal.

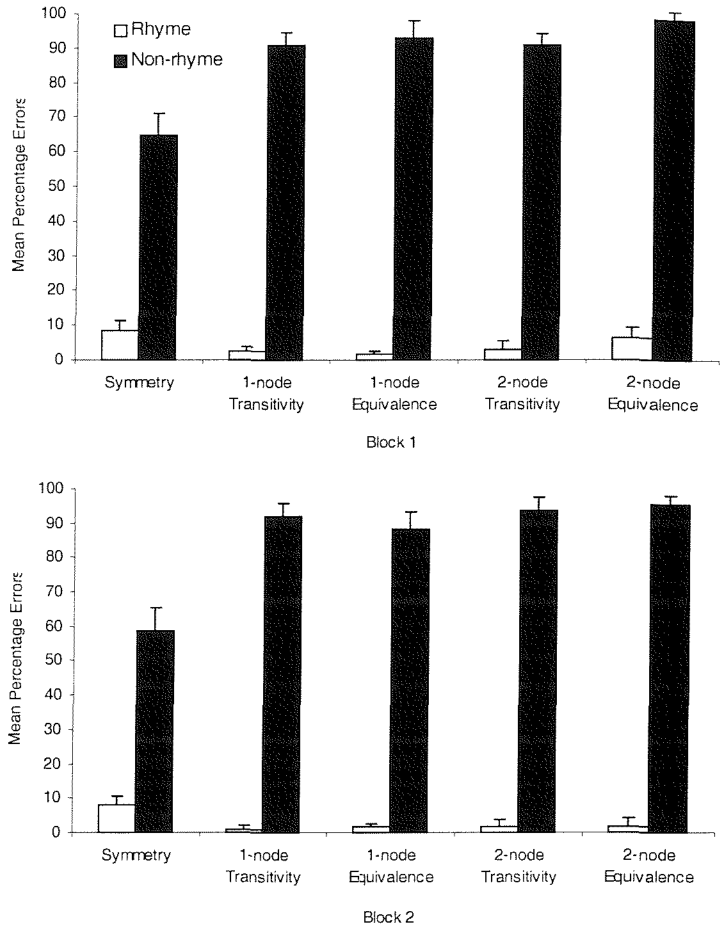

Figure 20. Means of percentage error scores (+*SE*) of all participants during testing of emergent relations (Phase 8).

Participants' mean response latencies were also subjected to a within-participant analysis of variance, which indicated significant main effects of phonology, $F(1, 15) = 20.88, p < .0001$, and of relation type, $F(4, 60) = 7.73, p < .0001$. [3] Participants' mean latencies were significantly greater on non-rhyme trials. Post hoc means comparisons (at an alpha level of .05) further indicated that participants' mean latencies were significantly greater on rhyme trials of two-node equivalence than on rhyme trials of any other relation type, and that mean latencies were significantly shorter on rhyme trials of symmetry than on rhyme trials of one- or two-node transitivity. Mean latencies on non-rhyme trials of two-node equivalence were also significantly greater than on non-rhyme trials of any other relation type except one-node equivalence. The top half of Figure 21 shows

---

[3] Because of the small proportion of correct responses on trials involving non-rhyme classes, analyses were performed on latencies for both correct and incorrect responses throughout.

participants' mean response latencies during this block for each type of emergent relation tested.

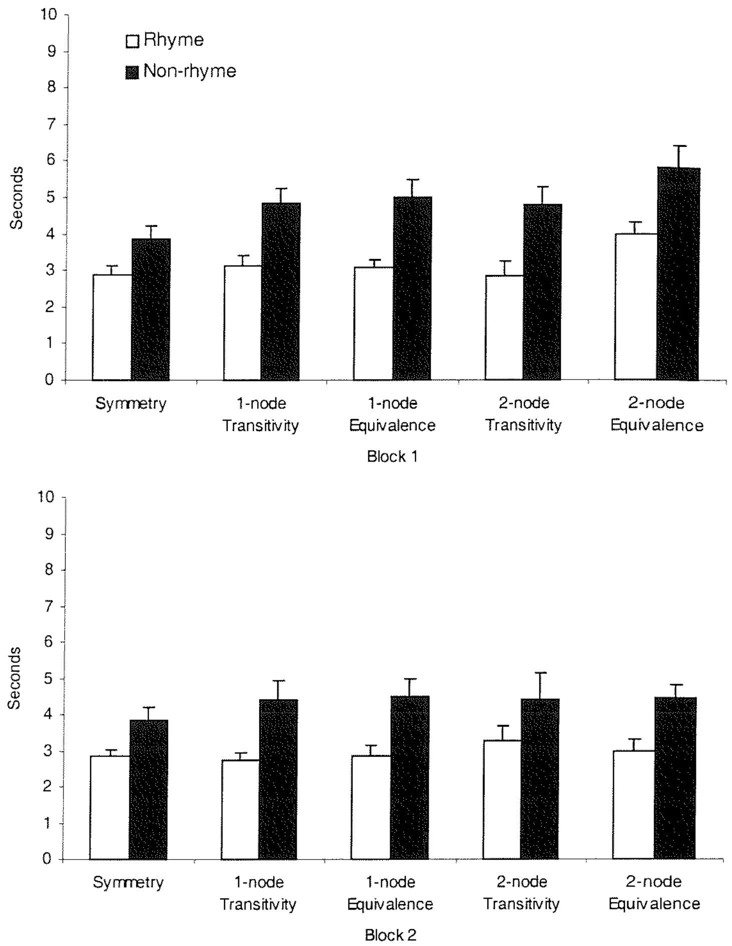

Figure 21. Mean response latencies (+*SE*) of all participants during emergent testing (Phase 8).

5.3.3.3 Review prior to final emergent testing.

Six participants (AA1, MS, RR, RW, AW, and AA2) demonstrated full baseline maintenance to criterion in extinction immediately following the first block of emergent testing, and two others (JW1 and CB) made only three errors each. Of the 10 participants who had failed to meet criterion during the first extinction full baseline review block, none required more than one reinforcement full baseline review block to re-establish baseline relations to criterion except AC and CB, who required two blocks. Participants made, on average, 1.8 (*SD* = 1.7) and 1.5 errors (*SD* = .7) during the first and second reinforcement full baseline review blocks, respectively. All participants satisfied the criterion for full baseline review during the second extinction block and proceeded to emergent testing. The right sections of Figure 18 shows total errors made by all participants during both blocks of extinction full baseline review prior to final emergent testing.

5.3.3.4 Final emergent testing (block 2).

Although all participants demonstrated baseline maintenance to criterion in extinction immediately prior to final emergent testing, again none demonstrated the emergent relations indicative of the formation of contextually controlled equivalence classes. Although participants made on average 31.3 errors ($SD = 5.2$) during the second block of emergent testing, on the 36 rhyme trials they made on average only 1.3 errors ($SD = 1.8$). Again, only two participants (AT and AH) made more than three errors on these trials, and 11 made one or less. On average, 1.1 ($SD = 1.8$) of these errors were other-context selections. On the 36 non-rhyme trials, participants made an average of 30 errors ($SD = 5.1$), and nine participants made 31 or more. On average, 26.3 ($SD = 9.6$) of these errors were other-context selections. The lower half of Figure 19 shows each participant's errors on non-rhyme trials, and the number of other-context selections on those trials.

Participants' percentage error scores were subjected to a within-participant analysis of variance in which the factors were again phonology and relation type. There was a significant main effect of phonology, $F(1, 15) = 791.59, p < .0001$, and of relation type, $F(4, 60) = 7.71, p < .0001$, and a significant interaction of Phonology x Relation Type, $F(4, 60) = 14.43, p < .0001$. Participants made significantly fewer errors on rhyme trials. Post hoc means comparisons (at an alpha level of .05) again indicated both that participants made significantly fewer errors on non-rhyme symmetry trials than on non-rhyme trials of any other relation type, and that there were no significant differences in participants' error scores on rhyme trials. The lower half of Figure 20 shows means of percentage error scores during the final block of emergent testing.

Participants' mean response latencies were also subjected to a within-participant analysis of variance in which the factors were again phonology and relation type. There was a significant main effect of phonology, $F(1, 15) = 13.39, p < .01$. Participants' response latencies were significantly greater on non-rhyme trials. The lower half of Figure 21 shows participants' mean response latencies during this block for each type of emergent relation tested.

5.3.4 Phase 9: Initial Generalisation Testing

Nine participants selected only comparisons whose names rhymed with those of the samples in the presence of both contextual stimuli, and three others selected two non-rhyme comparisons or less. Participant AW selected only rhyme comparisons in the presence of the rhyme contextual stimulus, however, and 17 of 18 non-rhyme comparisons in the presence of the non-rhyme contextual stimulus. None of the three remaining participants selected fewer than 16 rhyme comparisons on generalised rhyme trials, and only one of these selected fewer than 15 non-rhyme comparisons on generalised non-rhyme trials. On average, participants made .4 non-rhyme selections (*SD* = .8) on the 18 generalised rhyme trials (i.e., trials presenting stimuli in the presence of the rhyme contextual stimulus), and 3.7 non-rhyme selections on the 18 generalised non-rhyme trials (i.e., trials presenting

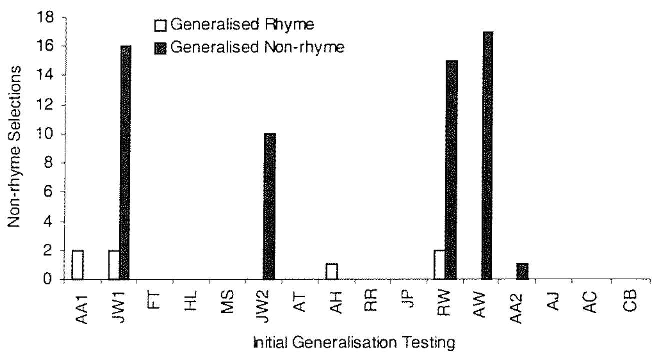

Figure 22. All participants' non-rhyme selections on generalised rhyme and non-rhyme trials during initial generalisation testing (Phase 9).

stimuli in the presence of the non-rhyme contextual stimulus). Non-rhyme selections were defined as selection of either of the comparisons whose names did not rhyme with that of the sample. Figure 22 shows all participants' non-rhyme selections on generalised rhyme and non-rhyme trials during initial generalisation testing.

5.3.5 Phase 10: Secondary Generalisation Testing

Three participants (JW1, RW, and AA2) met criterion for differential selection of rhyme and non-rhyme keys during the first block of secondary generalisation testing. On average, participants made 1.9 non-rhyme key selections (*SD* = 2.1) on the six trials presenting a rhyme compound as sample (rhyme compound trials), and an average of 3.8 non-rhyme key selections (*SD* = 1.5) on the six trials presenting a non-rhyme compound as sample (non-rhyme compound trials). Of the 13 participants who did not meet criterion during the first block of secondary generalisation testing, a further three (FT, AH, and RR) met criterion during the second block. Figure 23 shows each participant's non-rhyme key selections in response to rhyme and non-rhyme compounds during secondary generalisation testing.

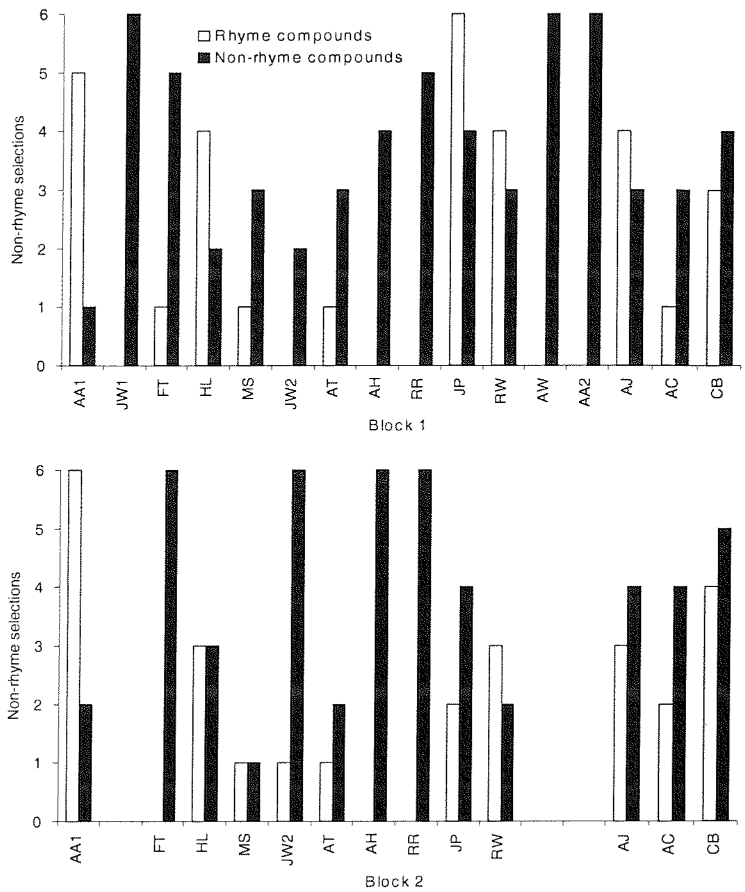

Figure 23. All participants' non-rhyme key selections in response to rhyme and non-rhyme compounds during secondary generalisation testing (Phase 10). Absence of initials indicates that trials were not presented to that participant.

5.3.6 Phase 11: Naming Post-test

Participants' written post-tests suggested almost ubiquitously normative stimulus naming, although participant AA2 indicated that he had used normative and non-normative names for the same stimuli on rhyme and non-rhyme trials respectively (i.e., *flea/bug*, *rat/mouse*, *goat/beast*, and *hair/girl*) and that he had likewise referred to *bee* as *bug* or *flea*, but had never referred to that stimulus by its normative name. Participant FT also indicated that she had initially referred to *flea* as *bug*.

All participants' post-tests indicated that comparisons had been selected on the basis of rhyme on rhyme trials during baseline training and emergent testing, but that on non-rhyme trials a variety of strategies had been employed. Four participants (JW1, HL, RR, and MS) indicated that they had remembered pairs of stimuli on these trials, although HL additionally noted selecting comparisons on the basis of rhyme when its paired stimulus was unavailable on non-rhyme trials (i.e., on trials of transitivity and equivalence during emergent testing). Five other participants (AA1, AA2, AH, AW, and AC) indicated selecting stimuli likewise, although AC also noted using intraverbal phrases (e.g., *the tree has a cat in it*, *bats have fleas*) on non-rhyme trials. Participant RR also indicated using intraverbal phrases to link stimuli, as did six other participants (JW2, JP, AW, CB, AH, and AA2). Two of these participants (JW2 and JP) also noted using mental images to link stimuli, as did participants RW and AJ. Participant JW2 further indicated using idiosyncratic spatio-temporal strategies, as did AH. Participant FT noted selecting comparisons on the basis of "indirect rhyme" (e.g., *tree-cat*, *bee-hat*) and idiosyncratic associations (e.g., *key/note* = *musical terms*) on non-rhyme trials.

Three participants (JW1, AW, and RW) indicated selecting rhyme and non-rhyme comparisons on rhyme and non-rhyme trials respectively, although the latter participant noted having made errors. Participants JW2 and AH indicated that they had selected rhyme comparisons on rhyme trials but that on non-rhyme trials they had again selected comparisons on the basis of idiosyncratic spatio-temporal strategies. Nine participants (AA1, FT, HL, MS, RR, JP, AA2, AJ, and CB) indicated that they had selected comparisons on the basis of rhyme throughout initial generalisation testing, regardless of the contextual stimulus.

Regarding secondary generalisation testing, six participants (JW1, FT, JW2, AH, RR, AW, and AA2) indicated that they had selected rhyme and non-rhyme keys in response to rhyme and non-rhyme compounds respectively. Participant AA1 indicated selecting non-rhyme keys in response to rhyme compounds and rhyme keys in response to non-rhyme compounds as did participant JP, who also noted that she had subsequently reversed her strategy on the basis of previous training. Participants RW and AJ indicated selecting comparisons on the basis of idiosyncratic strategies (e.g., *moon* = *night* = *blue*, *padlock* = *danger* = *red*) and participant HL indicated that she "did not name stimuli" and so had "used no strategies".

5.4 DISCUSSION

The results supported those of Study One, again suggesting that participants named the experimental stimuli without instruction, and that the phonological properties of the names thus given had influenced match-to-sample performance. Baseline establishment (phases 1 and 4) consisted of six training blocks (AB, BC, CD, EF, FG, and GH), and each block presented novel sample-correct comparison pairings. Three blocks, presented consecutively in the presence of the rhyme contextual stimulus, trained relations between stimuli whose names rhymed (rhyme blocks) and the other three blocks, also presented consecutively but in the presence of the non-rhyme contextual stimulus, trained relations between stimuli whose names did not rhyme (non-rhyme blocks). Although only two participants made more than a single error during the second rhyme block and only one participant made any error during the third block, by contrast, only one participant performed without error during either the second or third non-rhyme blocks.

Because the trial configurations used to establish baseline relations were identical for rhyme and non-rhyme classes, it seems plausible to suggest that the consistently class-congruent selection of comparisons observed during establishment of rhyme baseline relations was verbally controlled (Horne & Lowe, 1996) or rule-governed (Skinner, 1969), in that participants learned that selection of any comparison whose name rhymed with that of the sample would be correct in the presence of the rhyme contextual stimulus. In the presence of the non-rhyme contextual stimulus, however, participants learned a series of purely arbitrary visual discriminations, and had no such straightforward verbal basis for selection available. The virtually errorless maintenance of rhyme baseline relations demonstrated by participants during reinforcement and extinction rhyme baseline reviews (phases 2 and 3, or 5 and 6, respectively) did not undermine this interpretation.

Despite having demonstrated mastery of component rhyme and non-rhyme baseline relations, only four of 16 participants met criterion during their first block of extinction full baseline review (Phase 7). Effects of baseline training order were implicated by the pattern of errors observed during that block: Of a total of 88 errors made by participants who had established non-rhyme baseline relations immediately prior to full baseline review, only one error was made on non-rhyme trials. Likewise, of a total of 41 errors made by participants who had established rhyme baseline relations immediately prior to full baseline review, only one error was made on rhyme trials. Of the 129 errors made during this block, 125 were other-context selections. No significant effects of training order were observed in participants' performance during any subsequent training or testing block, however, and it would appear that participants' behaviour during the first full baseline review block subsequent to initial emergent testing was again the result of verbal control. Only eight of 78 errors made by all participants in this block were on rhyme trials.

It also seems plausible to suggest that participants' behaviour during emergent testing (Phase 8) was a product of the same verbal control: Although all participants had met criterion for full baseline review immediately prior to both initial and secondary emergent testing, none demonstrated the formation of contextually controlled equivalence classes in either block. During the first block of emergent testing, however, only two participants made more than three errors on the 36 rhyme trials and, of a total of 27 errors made by all participants during these trials, 17 were other-context selections. By contrast, only six participants made fewer than 30 errors on the 36 non-rhyme trials. Of a total of 465 errors made during this block, 380 were other-context selections. During both emergent testing blocks, no significant effects of nodal distance were observed in the accuracy with which participants responded on rhyme trials. In accordance with previous research (Kennedy, 1991; Kennedy et al., 1994), however, participants performed with greater accuracy on non-rhyme symmetry trials than on non-rhyme trials of any other relation type. That significantly greater latencies were observed on rhyme trials of two-node equivalence than on rhyme trials of any other relation type also supported previous reports of the transitivity latency effect (Dickins et al., 1993).

Although initial generalisation testing (Phase 9) presented only novel stimuli as samples and comparisons, nine participants selected rhyme comparisons throughout, and three others selected two non-rhyme comparisons or less. The performance of the four other participants, however, suggested that the contextual cues established during baseline training had controlled comparison selection. Because the stimuli involved bore no formal resemblance to each other or to the stimuli used in baseline training and emergent testing, and their names were also phonologically unrelated to the names of the baseline stimuli, it again seems plausible to suggest that that the contextual control observed was verbal in nature.

Secondary generalisation testing presented multi-element stimuli as samples, each composed of novel pairs of stimuli whose names either rhymed or did not rhyme with each other. Two comparisons were presented on every trial (i.e., rhyme and non-rhyme keys), each of the same colour as one of the contextual stimuli that had previously controlled selection of either rhyme or non-rhyme stimuli. Again, the pictorial stimuli employed bore no formal resemblances to each other or to the stimuli used in baseline training and emergent testing, and their names were also phonologically unrelated to the names of those stimuli. Nevertheless, six participants selected only rhyme keys in response to rhyme compounds and non-rhyme keys in response to non-rhyme compounds, although only three did so during the first testing block. Additional confirmation that stimuli were named was provided by written-post-testing, which also indicated that participants' stimulus naming had been almost ubiquitously normative. All participants reported that they had selected comparisons on the basis of rhyme in the presence of the rhyme contextual stimulus throughout baseline training and testing and emergent testing, but that on non-rhyme trials they had employing one or more of a variety of strategies, verbal and otherwise, to link stimuli.

In summary, therefore, the results of Study Two provide a striking demonstration of the functionality of participants verbal behaviour during equivalence research. Additional findings are indicated, however: Firstly, that previously established contextual control may be superseded by verbal control during testing without reinforcement if a ready verbal basis for categorisation of stimuli is available. Secondly, participants' performance during generalisation testing indicates that previously established contextual control can generalise to control differential selection of novel stimuli on the basis of participants' verbal behaviour. In addition, the results indicated that the contextual stimuli exerting that control can enter into membership of generalised classes composed of novel multi-element stimuli on the basis of the phonological characteristics of the names of the elements composing those stimuli.

6. STUDY THREE: NAMING AND CONTEXTUAL CONTROL--A PROCEDURAL REFINEMENT

6.1 INTRODUCTION

The results of Study Two again indicated that participants had named the experimental stimuli without instruction, that the phonological properties of the names thus given had influenced match-to-sample performance, and that baseline training may be superseded by verbal control during testing without reinforcement if a ready verbal basis for stimulus classification is available: Despite mastery and maintenance of contextually controlled baseline relations, participants majoritatively selected only rhyme comparisons throughout emergent testing, regardless of the contextual cues presented. Previously established contextual cues were observed to have controlled selection of novel stimuli, however, and to have entered into membership of generalised classes composed of novel stimulus compounds on the basis of participants' naming. Although a striking demonstration of the functionality of participants' verbal behaviour had thus been achieved, the principal aim of Study Two had nevertheless been to investigate the functionality of that behaviour in the contextual control of equivalence classes--and that control had not been established: Could this failure have resulted from procedural considerations, however?

A number of studies have suggested relationships between equivalence class formation and human memory (e.g., Green et al., 1990; Mackay, 1991; Mackay & Ratti, 1990; Saunders et al., 1988, 1990), and other research has indicated potential links between the number of stimulus relations with which participants are trained and the probability that those participants will demonstrate equivalence: As Spradlin et al. (1992) have pointed out, class-size may represent a variable that can affect both the formation and retention of equivalence classes--although definitive experiments on class-size limitations have yet to be carried out (Sidman, 1994). A substantial literature also attests to the delayed emergence of equivalence during repeated testing in extinction (e.g., Bush et al., 1989; Harrison & Green, 1990; Lazar et al., 1984; Sidman et al., 1985, 1986; Spradlin et al., 1973). In Study Two, a large number of stimulus relations had been trained and tested within a single experimental session during which only two emergent testing blocks had been presented. Class-size limitations therefore represented a possible procedural explanation for the non-emergence of equivalence between non-rhyme stimuli observed, as did the limited scope allowed for its potentially delayed emergence.

Study Three therefore set out again to investigate the functionality of participants' verbal behaviour in the contextual control of equivalence classes by refinement of certain elements of the procedure employed in Study Two. The number of stimuli composing each potential equivalence class was reduced from four to three, and the number of classes employed was reduced likewise. To accomplish this, certain of the stimuli that had been indicated by some participants to have been non-normatively named (e.g., *boat*, *rat*, *flea*, *note*, *hair*) during Study Two were not presented. To allow scope for the potentially delayed emergence of equivalence between non-rhyme stimuli, more extensive exposure to the experimental contingencies was also provided: The number of emergent blocks presented was quadrupled, and all participants were requested to attend two training and testing sessions on consecutive days, if required. The colours of the lines composing pictorial stimuli replaced background colours as contextual cues, and errorless training and testing criteria for baseline relations were also implemented (cf. Stikeleather & Sidman, 1990).

Further procedural extensions were also employed. Study Two had presented only one combination of non-rhyme stimuli to all participants (i.e., that presented to participants in the diagonal condition in Study One). This required that a comparison whose name rhymed with that of the sample had always been available to participants, and a ready verbal basis for stimulus classification (i.e., rhyme) had therefore been available on every trial. Because of this, it was decided additionally to employ another stimulus arrangement in which such comparison stimuli were not presented (i.e., a rearrangement of stimuli similar to that presented to participants in the orthogonal condition in Study One). A mixed design was therefore employed. In one condition, participants were presented with rhyme trials regardless of contextual cue. Participants in a second condition were presented with rhyme trials under the control of one contextual cue, and non-rhyme trials similar to those presented in Study Two under the control of the other contextual cue. Participants in a third condition were likewise presented with rhyme trials under the control of one contextual cue. Under the control of the other contextual cue, however, non-rhyme trials were presented involving stimuli in the class arrangements latterly described above. Participants in a fourth condition were presented with both of the aforementioned non-rhyme class arrangements, each under the control of one of the contextual cues.

Because of the suggestive yet inconclusive findings of generalisation testing in Study Two, more extensive testing was employed to assess the generalisation of contextual control to classes composed of novel stimuli on the basis of participants' potential naming of stimuli, and the number of stimuli composing potential classes during initial generalisation testing was reduced from three to two. As during Study Two, written post experimental tests of naming were conducted both as a measure of the normativeness of participants' stimulus naming and as an indicator of potential verbal strategies employed by participants during match-to-sample training and testing. Effects of nodal distance (Fields et al., 1990; Fields & Verhave, 1987) were also assessed.

6.2 METHOD

6.2.1 Participants

Twenty-four students and staff at the University of Southampton (10 female, 14 male) volunteered to participate in the study for a maximum of two experimental sessions and were assigned randomly, but in equal number, to four experimental conditions; rhyme/rhyme (R/R), rhyme/orthogonal (R/O), rhyme/diagonal (R/D), and diagonal/orthogonal (D/O). Aged between 18 and 33 years, all participants were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 per 30 min, independent of experimental performance.

6.2.2 Apparatus and Setting

Using software specifically designed for equivalence research (Dube & Hiris, 1996), a Power Macintosh® computer presented all stimuli and automatically recorded participants' responses and response latencies. During match-to-sample trials, its 15-in. (38-cm) monitor displayed five transparent keys (4.5 cm key) that were indiscernible against a white background. Sample stimuli were presented on the centre key, and comparisons appeared on the four outer keys. During baseline training and emergent testing, one of the outer keys, its position varying from trial to trial, always remained blank (see Figure 24). Throughout generalisation testing, two of the outer keys, their positions varying from trial to trial, always remained blank. Participants were tested individually in a small windowless cubicle (1.5 m by 2.9 m) containing a desk, on which were placed a sheet of written instructions, the computer, monitor, and mouse, and an envelope concealing a pen and post-test booklet for completion subsequent to match-to-sample testing. No keyboard was visible, and responses were made using the mouse. All participants completed the experiment in a maximum of two sittings, each on consecutive days (Days One and Two). Neither sitting ever exceeded 1.5 hr duration.

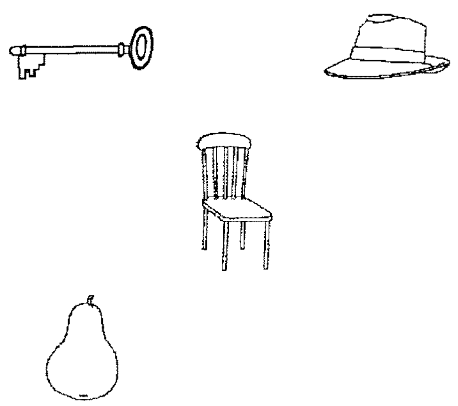

Figure 24. Typical match-to-sample screen display, illustrating an F2G2 trial for the R/D condition; *chair* (sample), *key* (correct comparison).

6.2.3 Stimuli and Class Arrangements

Baseline training and emergent testing			Initial generalisation testing			Secondary generalisation testing	
tree	bee	key	dog	frog	log	fan	man
bat	cat	hat	snake	cake	rake	well	bell
bear	chair	pear				spoon	moon

Table 10. Normative names of pictorial stimuli used in baseline training, emergent and initial generalisation testing, and elements of stimulus compounds used in secondary generalisation testing.

Throughout the experiment, pictorial stimuli were red, blue, or black line drawings of easily nameable items (some adapted from Snodgrass & Vanderwart, 1980), the normative names of which were each between three and five letters in length (see Table 10). During baseline training and emergent testing, stimuli were nine red and nine blue line drawings of the same items, three different arrangements of which each provided the potential for 3 three-member equivalence classes (three rhyme, three orthogonal, and three diagonal classes; shown in the top three sections of the right column of Figure 25, respectively). Rhyme classes were composed of stimuli

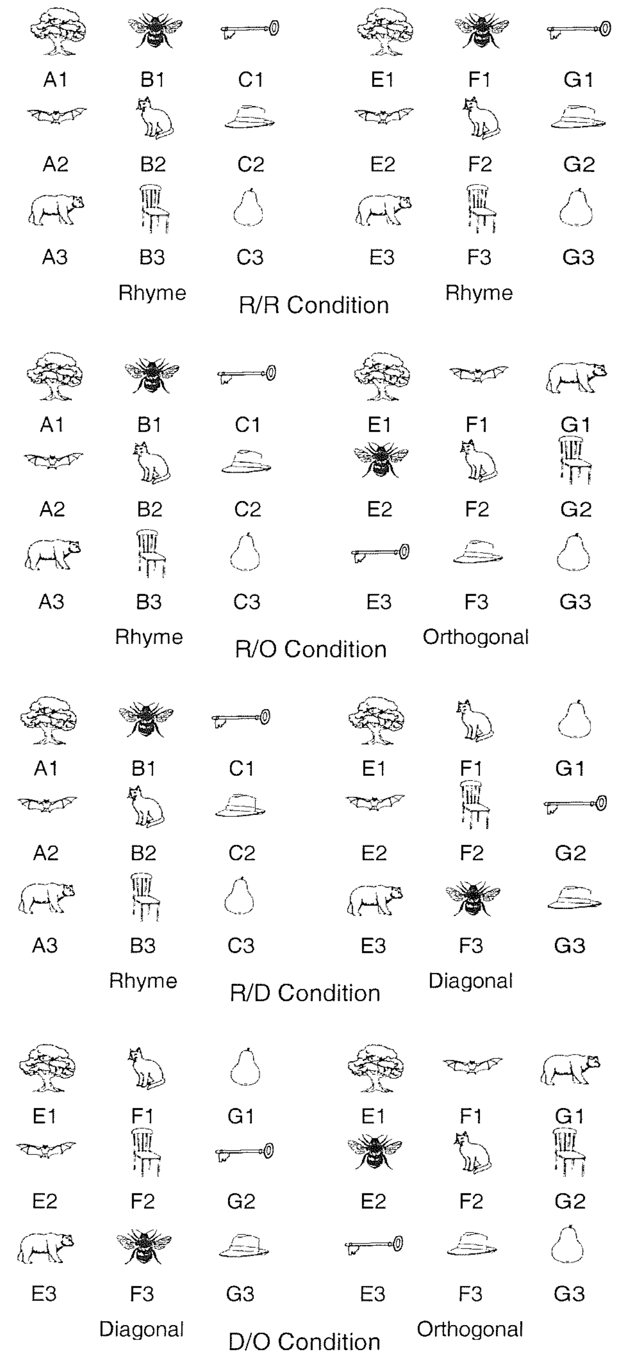

Figure 25. Stimuli and class configurations used in baseline and emergent trials for all conditions. Numbered rows denote classes; lettered columns denote stimuli.

whose names rhymed with each other, and trials involving these classes (rhyme trials) always presented a sample whose name rhymed with that of the positive comparison but never with those of the two negative comparisons. Orthogonal and diagonal classes were composed of stimuli whose names did not rhyme with each other: Trials involving the former classes (orthogonal trials) always presented three comparisons whose names rhymed with each other but never with that of the sample, and trials involving the latter classes (diagonal trials) always presented one incorrect comparison whose name rhymed with that of the sample, but never with that of either the correct or the other incorrect comparison. Participants in all conditions were exposed to two of these class arrangements, under the contextual control of the red or blue colour of the stimuli (red and blue classes, respectively). Participants in the R/R condition were trained with rhyme classes regardless of stimulus colour. Those in both the R/O and R/D conditions were trained with rhyme classes under the control of red, but were respectively trained with orthogonal and diagonal classes under the control of blue. Participants in the D/O condition were trained with diagonal classes under the control of red and orthogonal classes under the control of blue (see Figure 25 for stimuli and class arrangements for all conditions).

During initial generalisation testing, stimuli were red line drawings of another six items and six blue line drawings of those same items. These stimuli permitted use of only two comparisons per trial, the name of one of which always rhymed with that of the sample but never with that of the other comparison (see Figure 26), and provided the potential for four contextually controlled three-member generalised classes.

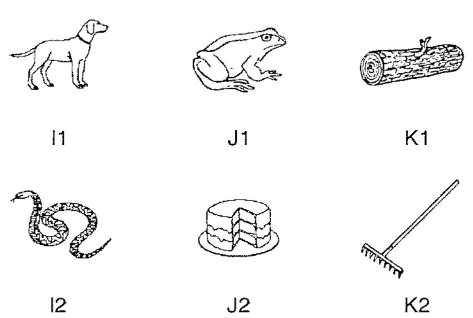

Figure 26. Stimuli used in initial generalisation testing for all conditions. Numbered rows denote rhyme classes; lettered columns denote stimuli.

During secondary generalisation testing, stimuli were a red and a blue key (of the same colours as the lines composing the pictorial stimuli described above) and six black line drawings, arranged into two classes of three stimulus compounds (see Figure 27). One of these classes was composed of pairs of stimuli whose names rhymed with each other (rhyme compounds) and the other was composed of pairs of stimuli whose names did not rhyme with each other (non-rhyme compounds). Throughout initial and secondary generalisation testing, stimuli were presented in identical class configurations for all conditions.

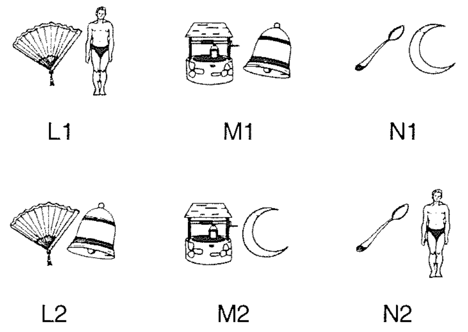

Figure 27. Stimulus compounds used in secondary generalisation testing for all conditions. Numbered rows denote classes; lettered columns denote stimulus compounds.

6.2.4 Procedure

Participants were asked to familiarise themselves with the instructions before them, and were then left to complete the experiment.

6.2.4.1 Instructions.

Initially, the following text was displayed on the computer's monitor: "When you are familiar with the written instructions, please click on 'Continue' to start the experiment." The written instructions were as follows:

> When the experiment begins, and at the start of each subsequent trial, you will see a picture in the middle of the screen in front of you. Use the mouse to click on it. More pictures will now appear in the corners of the screen. Use the mouse to click on one of these. At first, you will receive feedback on your choices, a "beep" for correct, and a "buzz" for incorrect. During later stages of the experiment, you will no longer receive feedback on your choices--the computer will tell you when. Keep on going however, and continue to do the best you can! Please aim to complete the experiment as quickly and accurately as possible. The computer will record your performance throughout, and a message on screen will tell you when the experiment is over. When you are ready to start, please click on "Continue".

Thank you for participating in this experiment. You are free to leave at any point.

The specified action removed the on-screen instructions and match-to-sample training commenced.

6.2.4.2 General procedure and match-to-sample contingencies.

Each trial began with presentation of a sample stimulus, an observing response with the mouse causing comparison stimuli additionally to be displayed. All stimuli remained in view until selection of a comparison caused them immediately to disappear, followed, after a 1-s interval, by presentation of the next trial. Comparison selections made within 0.5 s of presentation had no such consequence, however, and all stimuli remained in view. There was no limit to trial duration. Positions of correct and incorrect comparisons varied pseudo-randomly from trial to trial and, throughout baseline training and emergent and initial generalisation testing, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3, B4). At no point did the location of the correct comparison remain constant for more than two consecutive trials, nor did the same sample stimulus appear for more than two trials consecutively.

All participants were exposed to the same performance-contingent training and testing programme. During reinforcement training, selection of class-consistent comparison stimuli was followed by a beep and the word "CORRECT" displayed on the screen. Other choices resulted in a buzz and a darkened screen. During testing without reinforcement, however, the only consequence of a response was the presentation of the next trial. To assess effects of nodal distance, training was sequential (i.e., AB-BC, EF-FG), allowing the potential emergence of 12 symmetric (BA, CB, FE, GF), six transitive (AC, EG), and six equivalence relations (CA, GE). The overall procedure for all conditions was designed as a series of successive training and testing blocks, the details of which are given below.

6.2.4.3 Phase 1: Establish red (AB, BC) baseline relations.

Initially, AB relations were trained, with each of the three relations (A1B1, A2B2, A3B3) presented in pseudo-random order once every three trials. When a criterion of 12 consecutive correct responses had been achieved, BC relations were trained in identical fashion.

6.2.4.4 Phase 2: Review red baseline relations, with feedback.

Consequent to fulfilment of the above criteria, the baseline relations established during Phase 1 were reviewed in 12-trial blocks, with AB and BC trials intermixed in pseudo-random order. Samples from the same class were never presented consecutively. On completion of one reinforcement trial block with 100% accuracy, the next review phase commenced, assessing red baseline maintenance in extinction.

6.2.4.5 Phase 3: Review red baseline relations, without feedback.

Trained Relations											
AB			BC			EF			FG		
Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
A1	B1	B2, B3	B1	C1	C2, C3	E1	F1	F2, F3	F1	G1	G2, G3
A2	B2	B1, B3	B2	C2	C1, C3	E2	F2	F1, F3	F2	G2	G1, G3
A3	B3	B1, B2	B3	C3	C1, C2	E3	F3	F1, F2	F3	G3	G1, G2

Table 11. Baseline trial configurations, using single-sample and four comparison displays.

Otherwise identical to Phase 2, all trials during this phase were completed in the absence of explicit reinforcement (test trials). If performance remained at 100% accuracy over the first 12-trial block, training of EF and FG (blue) baseline relations commenced (Phase 4). If criterion was not achieved by the end of the second block, however, red baseline relations were reviewed, again with feedback (Phase 2). Review of red baseline relations and their maintenance in extinction continued in this way until 100% accuracy was demonstrated over one block of test trials. The left sections of Table 11 show the conditional discriminations established during phases 1 to 3 for all conditions.

6.2.4.6 Phases 4, 5, and 6: Establish and review blue baseline relations.

Phase 4 trained EF and FG baseline relations in exactly the same way that AB and BC relations were established during Phase 1. Likewise, phases 5 and 6 reviewed blue relations in the same way that red baseline relations were reviewed during phases 2 and 3. Subsequent to fulfilment of these criteria, the next review phase commenced, assessing maintenance of all red and blue baseline relations in extinction. The right sections of Table 11 show the conditional discriminations established during phases 4 to 6.

6.2.4.7 Phase 7: Full review of baseline relations.

Initially, all baseline relations established during phases 1 to 6 were reviewed in extinction. All AB, BC, EF, and FG trials were presented intermixed in pseudo-random order in 12-trial blocks. No more than two trials of red or blue baselines were ever presented consecutively. If participants satisfied the criterion of 12 of 12 correct responses, emergent testing commenced (Phase 8). If criterion was not met in this block, however, baseline relations were again reviewed in the same manner, but with feedback. On completion of one block of reinforcement trials to the same criterion, maintenance of the baseline relations was again assessed in extinction. Review of baseline and its maintenance continued in this way until criterion was met in extinction.

6.2.4.8 Phase 8: Emergent testing.

All possible emergent relations except reflexivity were presented, in pseudo-random order, in a maximum of four 24-trial blocks (see Table 12). Red and blue trials were never presented more than twice consecutively and generalised identity-matching repertoires were assumed. If participants satisfied the criterion of a minimum of 23 of 24 class-consistent responses in any one emergent testing block, initial generalisation testing commenced (Phase 9). If criterion had not been achieved at the end of the first two consecutive blocks of emergent testing, however, baseline relations were again reviewed, first without feedback (Phase 7) and, if criterion had not been achieved at the end of one trial block, again with feedback. Review of baseline continued in this way until all relations were again demonstrated to criterion over one block of test trials. The second two consecutive blocks of emergent testing were then presented. If criterion had not been achieved at the end of these blocks, match-to-sample testing ended for the day. An on-screen message automatically informed participants of this and reminded them to return for further testing on the following day. The general procedure for Day Two was identical to that for Day One, except that phases 1 to 6 were not presented and initial generalisation testing followed subsequent to completion of the fourth block of emergent testing, regardless of participants' performance.

	Emergent relations								
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
Symmetry									
BA	B1	A1	A2, A3	B2	A2	A1, A3	B3	A3	A1, A2
CB	C1	B1	B2, B3	C2	B2	B1, B3	C3	B3	B1, B2
FE	F1	E1	E2, E3	F2	E2	E1, E3	F3	E3	E1, E2
GF	G1	F1	F2, F3	G2	F2	F1, F3	G3	F3	F1, F2
Transitivity									
AC	A1	C1	C2, C3	A2	C2	C1, C3	A3	C3	C1, C2
EG	E1	G1	G2, G3	E2	G2	G1, G3	E3	G3	G1, G2
Equivalence									
CA	C1	A1	A2, A3	C2	A2	A1, A3	C3	A3	A1, A2
GE	G1	E1	E2, E3	G2	E2	E1, E3	G3	E3	E1, E2

Table 12. Emergent trial configurations, using single-sample and three comparison displays.

6.2.4.9 Phase 9: Initial generalisation testing.

A maximum of two blocks of 24 test trials each were presented, involving previously unseen stimuli. Each trial presented a novel sample stimulus followed, after an observing response, by two novel comparisons (see Table 13). In both blocks, 12 trials presented stimuli in the presence of the red contextual stimulus, and 12 presented the same stimuli in the presence of the blue contextual stimulus. All trials were intermixed in pseudo-random order and the same contextual stimulus was never presented on more than two consecutive trials. If participants satisfied a criterion of 23 of 24 class-consistent responses in the first block, secondary generalisation testing commenced. For participants in the R/R and D/O conditions, class-consistent responding was defined as selection of rhyme comparisons, regardless of contextual stimuli. For participants in the R/O and R/D conditions, class consistent responding was defined as selection of rhyme comparisons in the presence of the red contextual stimulus and selection of non-rhyme comparisons in the presence of the blue contextual stimulus. Secondary generalisation testing commenced subsequent to completion of the second block of initial generalisation testing, regardless of participants' performance.

Initial generalised relations (rhyme)						
	Sa	Co+	Co-	Sa	Co+	Co-
IJ	I1	J1	J2	I2	J2	J1
JI	J1	I1	I2	J2	I2	I1
JK	J1	K1	K2	J2	K2	K1
KJ	K1	J1	J2	K2	J2	J1
IK	I1	K1	K2	I2	K2	K1
KI	K1	I1	I2	K2	I2	I1

Initial generalised relations (non-rhyme)						
	Sa	Co+	Co-	Sa	Co+	Co-
IJ	I1	J2	J1	I2	J1	J2
JI	J1	I2	I1	J2	I1	I2
JK	J1	K2	K1	J2	K1	K2
KJ	K1	J2	J1	K2	J1	J2
IK	I1	K2	K1	I2	K1	K2
KI	K1	I2	I1	K2	I1	I2

Table 13. Initial generalisation testing trial configurations, using single-sample and two comparison displays.

6.2.4.10 Phase 10: Secondary generalisation testing.

Secondary generalised relations		
Sa	Co+	Co-
L1	rhyme key	non-rhyme key
M1	rhyme key	non-rhyme key
N1	rhyme key	non-rhyme key
L2	non-rhyme key	rhyme key
M2	non-rhyme key	rhyme key
N2	non-rhyme key	rhyme key

Table 14. Secondary generalisation testing trial configurations (blocks 1 and 4), using single-sample and two comparison displays.

Secondary generalisation testing was composed of six blocks of 12 test trials, each involving previously unseen stimuli (see Table 14). Blocks 1, 2, and 3 were identical to blocks 4, 5, and 6 respectively, and blocks were always presented in numerical order. Blocks 4, 5, and 6 were presented immediately subsequent to completion of Block 3, although presentation of those blocks was dependent upon participants' performance during blocks 1, 2, and 3, respectively (see below).

Blocks 1 and 4. Every trial presented a rhyme or non-rhyme compound as sample and both coloured keys as comparisons. Block 2 was presented subsequent to completion of Block 1, regardless of participants' performance. If participants selected only red keys in response to rhyme compounds and blue keys in response to non-rhyme compounds during Block 1, Block 4 was not presented.

Blocks 2 and 5. Every trial presented a red or a blue key as sample and a rhyme and non-rhyme compound as comparisons. Rhyme and non-rhyme comparisons were presented in quasi-random order. Block 3 was presented subsequent to completion of Block 2, regardless of participants' performance. If participants selected only rhyme compounds in response to the red key and non-rhyme compounds in response to the blue key during Block 2, Block 5 was not presented.

Blocks 3 and 6. Every trial presented a rhyme or non-rhyme compound as sample and a rhyme and a non-rhyme compound as comparisons. Rhyme and non-rhyme samples and comparisons were presented in quasi random order. If participants selected only rhyme compounds in response to rhyme compounds and non-rhyme compounds in response to non-rhyme compounds during Block 3, Block 6 was not presented.

When secondary generalisation testing had been completed, an on-screen message automatically informed participants that testing had ended and asked them to complete the post-test in the envelope before them.

6.2.4.11 Phase 11: Naming post-test.

Subsequent to match-to-sample testing, participants completed a written post-test that was designed to elucidate their verbal behaviour during the experiment. The booklet was headed by the following instructions:

> Printed below are the pictures that you have seen during the experiment. Did you mentally name any of them, or refer to them in any way during testing? If you did, please write under each picture the name, or names, you used for it during the experiment. If you did not refer to a picture in any such way, please leave the space underneath it blank.

Initially, the experimental stimuli used during baseline training and emergent testing were presented, each followed by a blank space and dotted line. Beneath this was a blank space headed by the following instructions:

> Please use the space below to describe briefly any mental strategies, rules, or other "tricks" that you may have used to learn and remember which of the above pictures went together. If the strategies you used differed depending on whether the pictures were RED or BLUE, please describe how they differed. If you did not use any such strategies, please leave the space blank and complete the rest of the post-test.

Following this, the experimental stimuli used during initial generalisation testing were presented, each followed by a blank space and dotted line. Above this were the following instructions:

> Later in the experiment, you saw the following pictures. As before, please note beneath each any name, or names, by which you may have referred to it during testing.

Next followed a blank space headed by the following instructions:

> In this later stage of the experiment, did you use any mental strategies, "tricks", or rules in choosing which pictures went together? If you did, please enter a brief description below, and as before indicate whether and how they may have differed when the pictures were RED and when they were BLUE. If you used no such strategies, please leave the space below blank and complete the rest of the post-test.

Lastly, the stimuli six elements of the stimulus compounds used during secondary generalisation testing were presented separately, each followed by a blank space and dotted line. Above these were the following instructions:

> In the final part of the experiment, you saw the following pictures. Once again, please note beneath each any name, or names, by which you may have referred to it during testing.

Beneath this were three blank spaces, above which were written "STAGE ONE", "STAGE TWO", and "STAGE THREE", respectively. Above these were the following instructions:

> In this final part of the experiment, there were three stages: In STAGE ONE, there were pictures in the middle of the screen, and you chose either a RED or a BLUE square on every trial. In STAGE TWO, there was either a RED or BLUE square in the middle of the screen, and pictures in two corners of the screen. In STAGE THREE, there were pictures only, and no coloured squares. Did you use any strategies, 'tricks', or rules to make your

choices? If you did, please enter a brief description in the space below. If not, leave the space blank.

6.3 RESULTS

All participants completed the experiment, although the accuracy with which they did so strongly differentiated participants in the R/R condition from those in the R/O, R/D, and D/O conditions, and participants in the D/O condition from those in the other conditions. Individual participants' trials and errors during all phases of match-to-sample training and testing are presented in Appendix E, and their latency data during emergent testing appear in Appendix F.

6.3.1 Phases 1 and 4: Establish Red and Blue Baseline Relations

Acquisition of red baseline relations was easier for participants in the R/O condition than for participants in the other conditions. The mean number of trials required by R/O participants to meet both criteria for this phase was 13.9 (*SD* = 2.9) with a mean error score of 1.2 (*SD* = 1.3), whereas the mean number of trials required by participants in the D/O condition was 23.6 (*SD* = 9.9), with a mean of 5.8 errors (*SD* = 5.1). Participants in the R/R condition required a mean of 14.3 trials (*SD* = 4.0) with a mean error score of 1.8 (*SD* = 2.9), and those in the R/D condition required a mean of 14.7 trials (*SD* = 5.0) with a mean error score of 1.8 (*SD* = 2.3). With only two exceptions, all participants in the R/R, R/O, and R/D conditions made fewer errors and required fewer trials to establish red baseline relations than any participant in the D/O condition. Only three participants in the former conditions required more than 28 trials to meet the criteria for red baseline relations, and only two made more than four errors each. Although the smallest number of trials required in these conditions was 24 (0 errors), no participant in the D/O condition required fewer than 33 trials (5 errors) to the meet the same criteria, and the greatest number of trials required was 60 (22 errors).

Regarding blue baseline relations, acquisition was easier for participants in the R/R condition than for participants in the other conditions. R/R participants required a mean of 12.1 trials (*SD* = 0.3) to meet both criteria for this phase with a mean error score of 0.1 (*SD* = 0.3), whereas participants in the R/D condition required a mean of 30.5 trials (*SD* = 13.5) with a mean error score of 8.2 (*SD* = 5.1). Participants in the R/O condition required a mean of 24.3 trials (*SD* = 6.4) with a mean error score of 6.6 (*SD* = 3.9), and those in the D/O condition required a mean of 20.7 trials (*SD* = 6.8), with a mean of error score of 3.8 (*SD* = 1.9). Only one participant in the R/R condition required more than 24 trials (0 errors) to establish blue baseline relations--the minimum to meet the criteria. No participant in the R/O, R/D, or D/O conditions required fewer than 28 trials to meet the same criteria, however, and the greatest number of trials required was 53. Although two participants in the latter conditions made only two errors each, six others each made 11 errors or more.

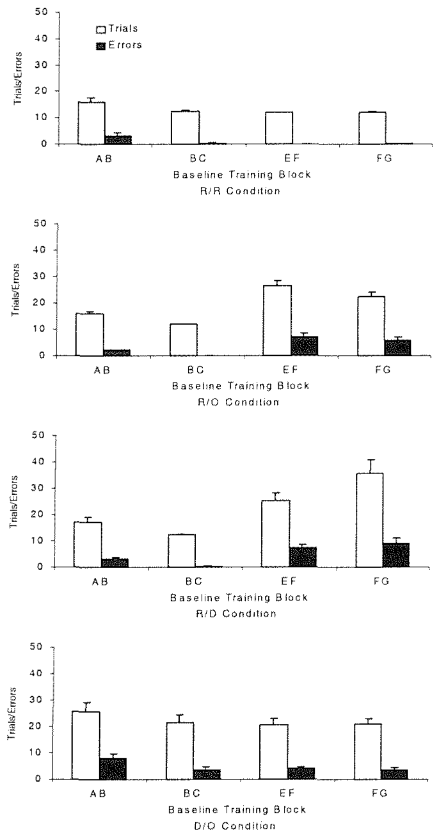

Figure 28. Mean trials and errors (+*SE*) for all participants in the R/R, R/O, R/D, and D/O conditions to establish red and blue baseline relations (phases 1 and 4).

The trial and error data were subjected to mixed design analyses of variance, in both of which the between-participant factor was condition (R/R, R/O, R/D, D/O) and the within-participant factor was context (red, blue). Regarding the number of trials required to meet the criteria for establishment of baseline relations, there was a significant main effect of condition, $F(3, 20) = 9.06$, $p < .01$, and a significant interaction of Condition x Context, $F(3, 20) = 5.04$, $p < .01$. Regarding errors, there was a significant main effect of condition, $F(3, 20) = 7.38$, $p < .01$, and a significant interaction of Condition x Context, $F(3, 20) = 5.74$, $p < .01$. Scheffé tests (at an alpha level of .05) indicated that although participants in the D/O condition required a significantly greater number of trials to meet the criteria than those in the R/R, R/O, or R/D conditions, the number of trials required by participants in the latter conditions did not differ significantly from each other. Although participants in the R/R condition made significantly fewer errors than participants in the D/O condition, no other significant differences were observed between error scores in the other conditions. Figure 28 shows the mean number of trials required and errors made by participants in all conditions during this phase of training.

6.3.2 Phases 2, 3, 5, and 6: Red and Blue Baseline Reviews

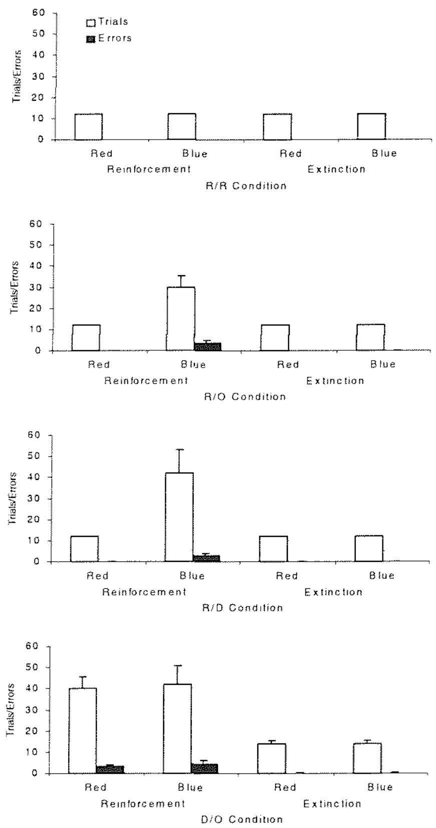

Figure 29. Mean trials and errors (+*SE*) for all participants in the R/R, R/O, R/D, and D/O conditions during reinforcement and extinction red and blue baseline reviews (phases 2, 3, 5, and 6).

All 18 participants in the R/R, R/O, and R/D conditions achieved complete accuracy during the first reinforcement review block of red baseline relations, but only one participant from the D/O condition performed likewise. Although all participants in the R/R condition again performed without error during the first reinforcement review block of blue baseline relations, no participant in the D/O condition and only two in the R/O condition and in the R/D condition performed likewise. The left sections of Figure 29 show the mean number of trials required and errors made by participants in all conditions during reinforcement review of red and blue baseline relations.

All participants performed without error during the first extinction review blocks of both red and blue baseline relations except D/O participants NM and LA, who made one error during red and blue extinction baseline reviews, respectively. The right sections of Figure 29 show the mean number of trials required and errors made by participants in all conditions during extinction review of red and blue baseline relations.

6.3.3 Phase 7: Full Baseline Reviews

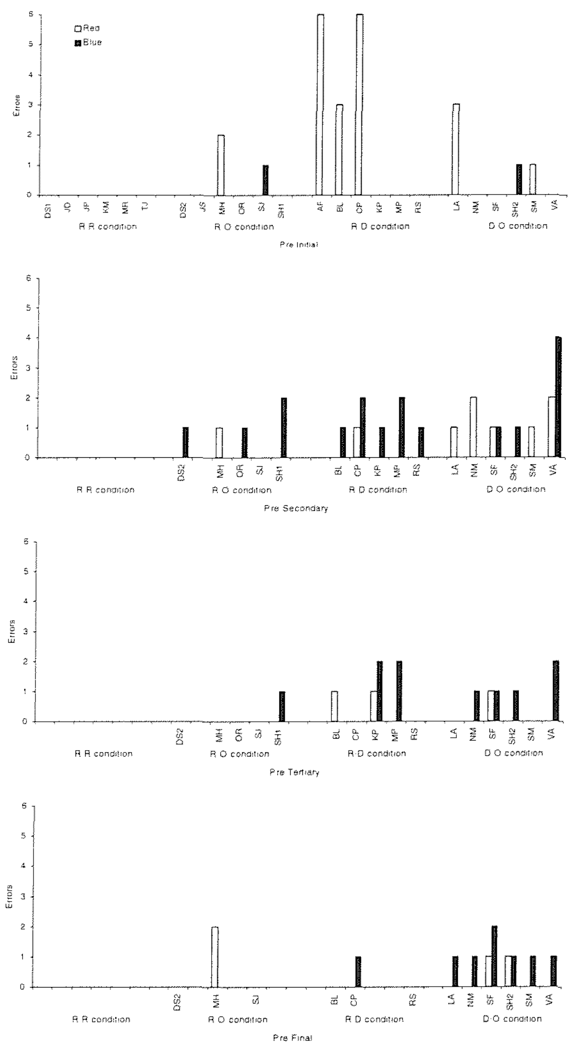

Figure 30. All participants' errors in the first block of extinction full baseline review (Phase 7) prior to initial, secondary, tertiary, and final emergent testing. Absence of initials indicates that trials were not presented to that participant.

Figure 30 shows the mean number of errors for every participant in all conditions during full baseline reviews prior to initial, secondary, tertiary, and final emergent testing.

6.3.3.1 Review prior to initial emergent testing.

All six participants in the R/R condition achieved complete accuracy during their first block of extinction full baseline review, but only four participants in the R/O condition and three in both the R/D and the D/O conditions performed likewise. Although three participants in the R/D condition and one participant in the R/O condition required only two additional reinforcement full baseline review blocks to re-establish baseline relations to criterion, one participant in the R/O condition, and all three remaining participants in the D/O condition required four blocks or more. All participants satisfied the criterion for extinction full baseline review during their second block except one participant in the D/O condition who required one further block, and one participant in the R/O condition, who required five further blocks.

6.3.3.2 Review prior to secondary emergent testing.

Because all participants in the R/R condition had met criterion for equivalence during initial emergent testing, none received further baseline review. Of the five R/O participants who had not met the criterion for equivalence during initial emergent testing, one demonstrated continued baseline maintenance in extinction and proceeded immediately to secondary emergent testing. No other participant in any of the conditions performed likewise, however, although none of these participants required more than three blocks of extinction full baseline review to re-establish baseline relations to criterion. Only one participant in the R/O and R/D conditions required more than six reinforcement full baseline review blocks to meet criterion during this review phase, but only two participants in the D/O condition required less than six blocks.

6.3.3.3 Review prior to tertiary emergent testing.

During the first block of extinction full baseline review (the first 12 trials of the second sitting), four of the five remaining participants in the R/O condition demonstrated continued baseline maintenance to criterion in extinction and proceeded immediately to tertiary emergent testing, as did two of the remaining five participants in the R/D condition and two of the six participants in the D/O condition. No other participant required more than one additional extinction full baseline review to re-establish baseline relations to criterion. Although no participant in the R/O or R/D conditions required more than three reinforcement full baseline review blocks to meet criterion during this phase, one participant in the D/O condition required four blocks and two others required 11 blocks.

6.3.3.4 Review prior to final emergent testing.

Of the three remaining participants in the R/O and R/D conditions, two in each condition demonstrated continued baseline maintenance in extinction immediately subsequent to tertiary emergent testing and proceeded to final emergent testing. The remaining participant in each condition required only one additional extinction full baseline review block and one reinforcement full baseline review block to re-establish baseline relations to criterion. All participants in the D/O condition required two extinction full baseline review blocks to re-establish baseline relations to criterion, except one who required three blocks. Although three participants in this condition required only one reinforcement full baseline review block to meet criterion during this review phase, the other three participants required 3, 12, and 16 blocks respectively.

6.3.4 Phase 8. Emergent Testing

Figure 31 shows means of percentage error scores made by participants in all conditions during initial, secondary, tertiary, and final emergent testing. Percentages have been presented because the numbers of emergent relations composing each block were unequal.

6.3.4.1 Initial emergent testing (blocks 1 and 2).

Although all participants had demonstrated full baseline maintenance to criterion in extinction immediately prior to initial emergent testing, equivalence was confined almost exclusively to participants in the R/R condition. Within the first block of emergent testing, all R/R participants but one satisfied the 23 of 24 criterion for equivalence, without error. One participant made two errors during the first testing block, but performed errorlessly during the second. By contrast, only one participant in the R/O condition and one in the R/D condition met criterion during the first two blocks of emergent testing, and no other participant in these conditions made fewer than four errors. In the D/O condition, no participant made fewer than six errors.

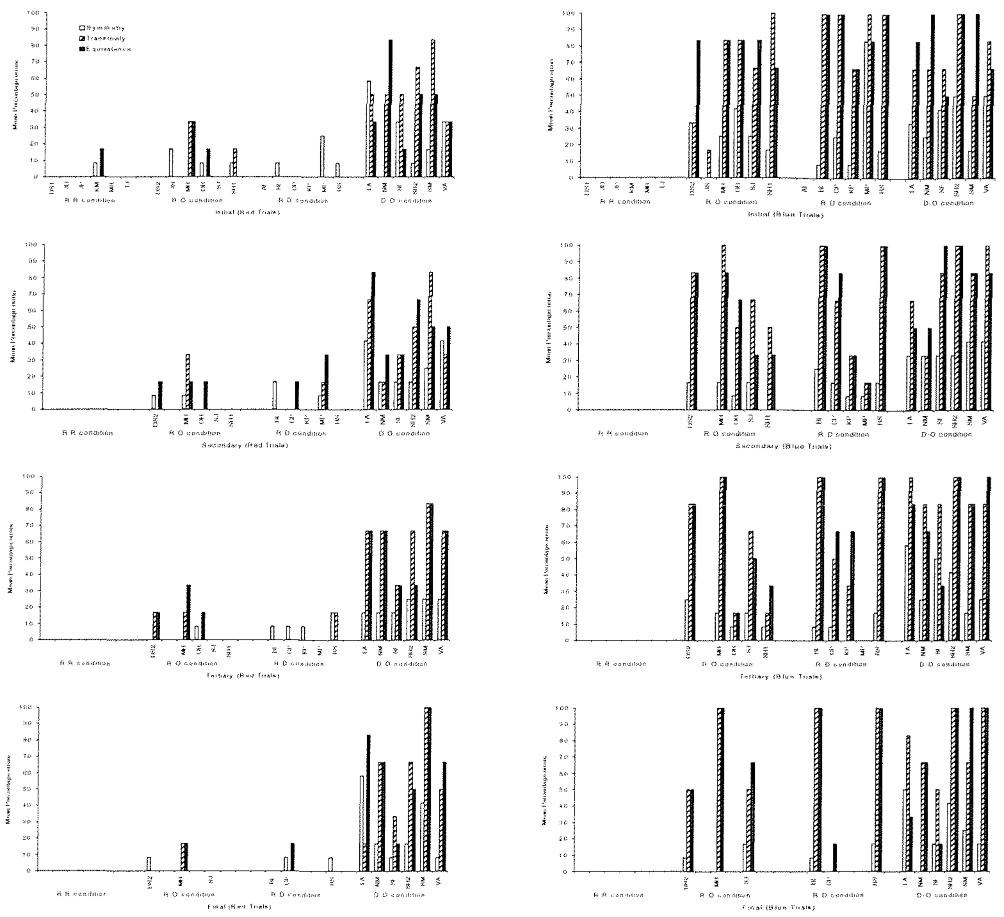

Figure 31. Percentage error scores of all participants in the R/R, R/O, R/D, and D/O conditions on red and blue trials during initial, secondary, tertiary, and final emergent testing (Phase 8). Absence of initials indicates that trials were not presented to that participant.

Participants' percentage error scores during the first block of emergent testing (the only block to which all participants were exposed) were subjected to a mixed design analysis of variance in which the between-participant factor was condition (R/R, R/O, R/D, D/O) and the within-participant factors were context (red, blue) and relation type (symmetry, transitivity, equivalence). There were significant main effects of condition, $F(1, 20) = 14.82, p < .001$, context, $F(1, 20) = 55.58, p < .001$, and of relation type, $F(2, 40) = 13, p < .001$. Significant interactions of Condition x Context, $F(3, 20) = 10.52, p < .001$, Context x Relation type, $F(2, 40) = 10.05, p < .001$, and Condition x Context x Relation Type, $F(6, 40) = 2.82, p < .05$ were also observed. Scheffé tests (at an alpha level of .05) indicated that although significantly fewer errors were made in the R/R condition than in the other conditions, error scores in those conditions did not differ significantly from each other. Post hoc means comparisons (at an alpha level of .05) further indicated that participants in the R/O, R/D, and D/O conditions made significantly fewer errors on blue trials of symmetry than on blue trials of transitivity or equivalence, and that participants in the D/O condition additionally made significantly fewer errors on red trials of symmetry than on red trials of transitivity. No other significant differences were observed.

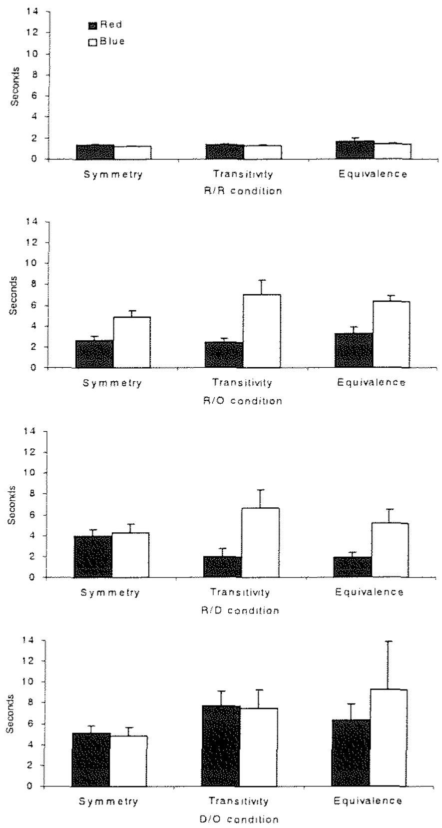

Figure 32. Mean response latencies (+*SE*) of all participants in the R/R, R/O, R/D, and D/O conditions during the first block of emergent testing (Phase 8).

Participants' mean response latencies during the first block of emergent testing were also subject to a mixed design analysis of variance, which indicated significant main effects of condition, $F(1, 20) = 6.75, p < .01$ and context, $F(1, 20) = 14.06, p < .01$, and a significant interaction of Condition x Context, $F(3, 20) = 3.38, p < .05$. Scheffé tests (at an alpha level of .05) further indicated that although latencies were significantly shorter in the R/R condition than in the D/O condition, latencies in those conditions did not differ significantly from latencies in the other conditions, whose latencies also did not differ significantly from each other. Post hoc means comparisons (at an alpha level of .05) indicated that the latencies of participants in the R/O and R/D conditions were significantly shorter on blue trials of symmetry than on blue trials of transitivity, and that the latencies of participants in the D/O condition were significantly shorter on blue trials of symmetry than on blue trials of equivalence. No other significant differences were observed. Figure 32 shows all participants' mean response latencies during the first block of emergent testing.

6.3.4.2 Secondary emergent testing (blocks 3 and 4).

No remaining participant met criterion for equivalence during this testing phase, although two participants in the R/O condition and one in the R/D condition made only two errors each during the fourth block of testing. Although one participant in the D/O condition made only three errors during that block, no other participant in that condition made fewer than 10 errors during secondary emergent testing.

6.3.4.3 Tertiary emergent testing (blocks 5 and 6).

One participant in the R/O condition and one participant in the R/D condition met criterion for equivalence during the fifth block of emergent testing--the latter performing without error. Although another participant from each of these conditions met criterion during the sixth testing block, no participant in the D/O condition made fewer than eight errors in either testing block.

6.3.4.4 Final emergent testing (blocks 7 and 8).

Of the remaining three participants in the R/D condition, one met criterion for equivalence during the seventh block of testing. No other participant performed likewise, however, although two participants in the R/O condition made only three and four errors respectively during that block. Only one participant in the D/O condition made fewer than ten errors during the seventh emergent testing block, although three made nine errors or less during the eighth block.

6.3.5 Phase 9. Initial Generalisation Testing

During the first 24-trial block of initial generalisation testing, all participants in the R/R condition fulfilled the 23 of 24 criterion for selection of rhyme comparisons without error, and proceeded immediately to secondary generalisation testing (Phase 10). No participant in the D/O condition performed likewise in either testing block, however, although two participants selected only non-rhyme comparisons during both testing blocks, as did a third during the second block only.

In the R/D condition, four participants satisfied the 23 of 24 criterion for differential selection of rhyme and non-rhyme comparisons during the first testing block, three performing without error. The remaining two participants made two and six errors respectively in both the first and second testing blocks, all but one of which were made on non-rhyme trials. In the R/O condition, only one participant met criterion during the first block, although two others made only two errors each during the second block: one of these participants performed without error during the second testing block. Although no other participant in this condition met criterion during the second block, two others made only two and three errors respectively. Of the remaining two participants, one selected only rhyme comparisons throughout both testing blocks, and the other did likewise during the second block. Figure 33 shows non-rhyme selections of every participant in all conditions during initial generalisation testing.

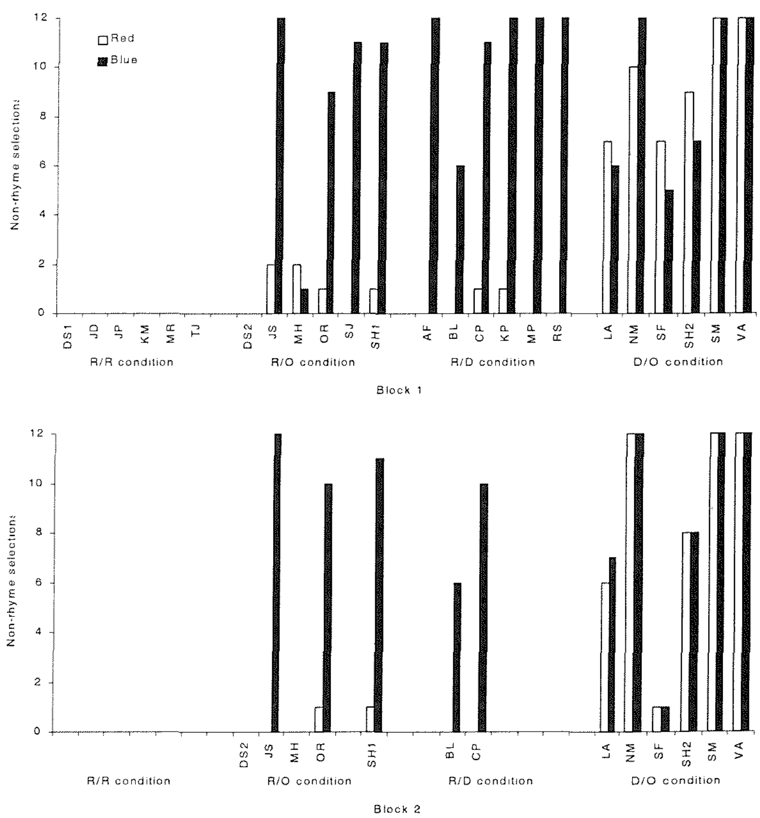

Figure 33. All participants' non-rhyme selections on red and blue trials during initial generalisation testing (Phase 9). Absence of initials indicates that trials were not presented to that participant.

6.3.6 Phase 10. Secondary Generalisation Testing

Figure 34 shows mean errors during all secondary generalisation testing blocks for participants in all conditions.

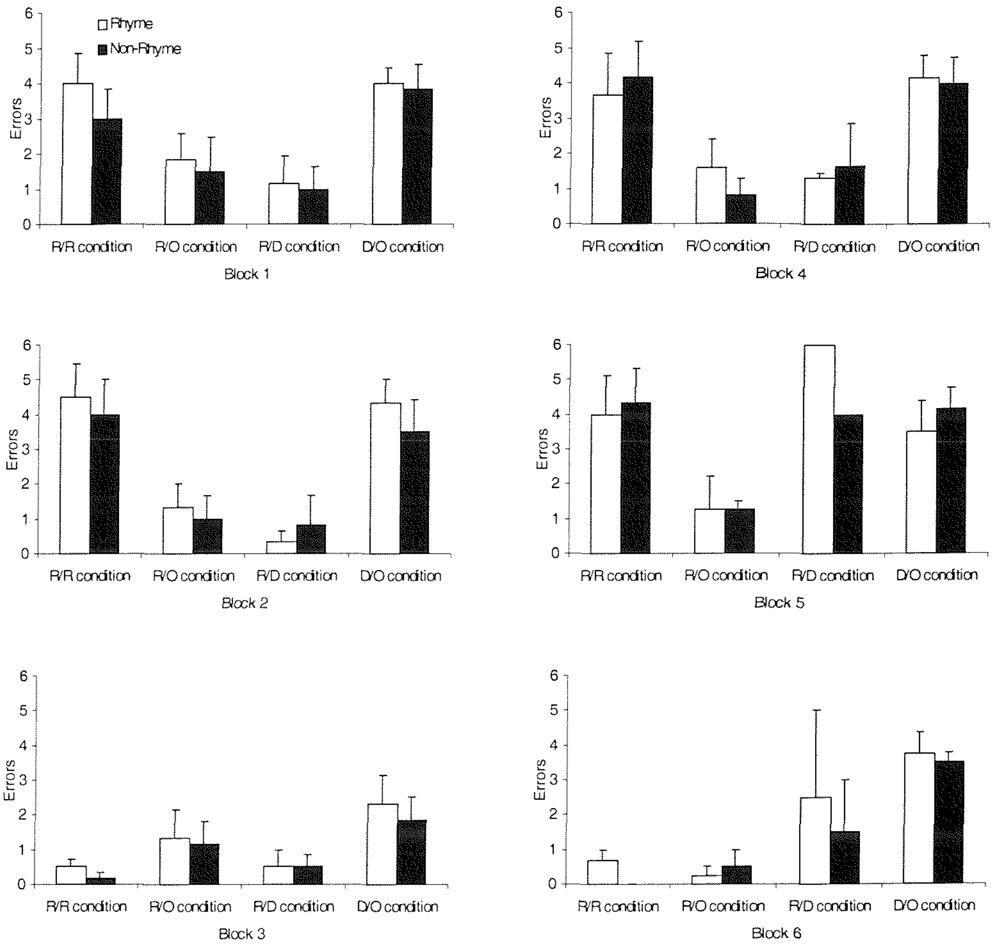

Figure 34. Mean errors (+*SE*) for all participants during secondary generalisation testing (Phase 10).

6.3.6.1 Blocks 1 and 4.

No participant in the R/R or D/O conditions met the errorless criterion for differential selection of red and blue comparisons in either testing block, although during Block 4, two participants in the former condition made only one error each. In the R/O condition, one participant performed without error during the Block 1, and two others met criterion during Block 4. In the R/D condition, three participants performed without error during Block 1, and another participant performed likewise during Block 4.

6.3.6.2 Blocks 2 and 5.

No participant in the R/R or D/O conditions met the errorless criterion for differential selection of rhyme and non-rhyme compounds in either testing block, although in the R/O condition two participants performed without error during Block 2 and another made only one error. In the R/D condition, all participants but one performed without error during Block 2.

6.3.6.3 Blocks 3 and 6.

Three participants in the R/R condition met the errorless criterion for differential selection of rhyme and non-rhyme compounds during Block 3, and another performed likewise during Block 6. In the R/O condition, two participants met criterion during Block 3, and another two performed likewise during Block 6. In the R/D condition, four participants met criterion during Block 3, and another performed likewise during Block 6. In the D/O condition, two participants met criterion during this phase of testing, both during Block 3.

6.3.7 Phase 11. Naming Post-test

Participants' post-tests suggested almost ubiquitously normative stimulus naming. All participants in the R/R and R/D conditions indicated using only normative stimulus names throughout the experiment, and only one participant in the R/O condition indicated otherwise (i.e., *fly* for *bee*). All participants in the D/O condition but one indicated using only normative stimulus names, although two noted that they had prefixed one or more of those names with another word (e.g., *polar bear*, *big key*, *Spanish fan*, etc.). One participant in the R/R condition and one in the R/D condition noted that they had spoken these names out loud throughout the experiment to facilitate comparison selection.

Regarding baseline training and emergent testing, no participant in the R/R condition indicated comparison selection on any basis other than that of the rhyme between stimulus names. In the R/D and R/O conditions, all participants likewise indicated comparison selection on the basis of rhyme in the presence of the red contextual stimulus, but that they had used a variety of strategies in the presence of the blue contextual stimulus: Five indicated using visual imagery on these trials (e.g., *bear eats the bee*, etc.), and four noted using verbal strategies (e.g., *chair* and *key* = *Cherokee*). Three participants indicated learning "sets" of stimuli (e.g., *tree* ⇔ *bat* ⇔ *bear*), and another indicated learning pairs of stimuli during baseline training, and four participants noted that they had chosen comparisons randomly, or on the basis of rhyme, during emergent testing. In the D/O condition, one participant indicated learning stimulus pairs dependent on the colour of the contextual stimulus, and four noted using verbal strategies--one indicating that he had incorporated the word "blue" in his naming of blue stimuli. Two other participants indicated using visual imagery.

Regarding initial generalisation testing, all R/R participants again indicated choosing only rhyme comparisons. In the R/D condition, five participants stated that they had selected rhyme comparisons in the presence of the red contextual stimulus. Although four of the latter participants indicated choosing non-rhyme compounds in the presence of the blue contextual stimulus, the other indicated choosing randomly. In the R/O condition, four participants indicated choosing rhyme and non-rhyme compounds in the presence of the red and blue contextual stimuli respectively, and two others indicated choosing only rhyme comparisons throughout. In the D/O condition, three participants indicated choosing only non-rhyme comparisons, and another indicated using a variety of idiosyncratic strategies (e.g., *dogs eat cakes*, etc.). The other two participants indicated that they had chosen comparisons randomly.

Regarding secondary generalisation testing, all R/R participants but one indicated that they had chosen red and blue keys differentially in response to rhyme and non-rhyme compounds during blocks 1 and 3. One indicated selection of the red key in response to the former samples, and four indicated selection of the blue key. The other participant indicated choosing the blue key regardless of the sample stimulus. In the R/D condition, all participants but one noted differential comparison selection, as did all participants but two in the R/O condition. Two participants noted using idiosyncratic strategies (e.g., *well = water = blue*), and one indicated choosing comparisons randomly. Two participants in the D/O condition noted differential comparison selection, three others noted using idiosyncratic strategies, and one indicated choosing randomly.

Regarding blocks 2 and 5, all R/R participants who had indicated that they had selected comparisons differentially during blocks 1 and 3 indicated that they had again performed in the same manner. Likewise, all participants in the R/D condition indicated using strategies similar to those they had indicated previously, as did all participants in the R/O condition except one, who noted changing from random to differential comparison selection. All D/O participants likewise indicated using strategies similar to those that they had indicated above.

All R/R participants indicated choosing rhyme compounds in response to rhyme compounds and non-rhyme compounds in response to non-rhyme compounds during blocks 3 and 6, as did all participants in the R/D condition except one, who noted using idiosyncratic strategies. All participants in the R/O condition indicated similar strategies except one, who noted using idiosyncratic strategies. In the D/O condition, only two participants indicated differential comparison selection. Three participants in this condition indicated using idiosyncratic strategies, and one noted choosing randomly.

6.4 DISCUSSION

The results of match-to-sample training and testing again suggested that participants named the experimental stimuli without instruction, and that the phonological properties of the names thus given had influenced match-to-sample performance. These findings were confirmed by written post-testing, which also confirmed that participants' naming of stimuli had been almost ubiquitously normative. Establishment of red baseline relations (Phase 1) consisted of two consecutively presented training blocks (AB and BC), each composed of newly introduced stimuli, yet all participants in the R/R, R/O, and R/D conditions but two performed without error during BC training. No participant in the D/O condition, however, performed without error during this block. Likewise, establishment of blue baseline relations (Phase 4) consisted of two consecutively presented training blocks (EF and FG). Although one participant in the R/R condition made a single error during FG training, no other participant in that condition made any error during either block. By contrast, no participant in the R/O, R/D, or D/O conditions performed without error during either of these blocks.

Because the baseline stimuli presented in all conditions bore no consistent formal resemblances to each other, it seems plausible to suggest that participants' consistently class-congruent selection of comparison stimuli on rhyme trials (i.e., trials of red baseline relations for participants in the R/R, R/O, and R/D conditions and also trials of blue baseline relations for participants in the former condition) was verbally controlled (Horne & Lowe, 1996), or rule-governed (Skinner, 1969), in that they had previously learned that selection of any comparison whose name rhymed with that of its sample would be correct. When presented with non-rhyme trials (i.e., trials of blue baseline relations for participants in the R/O, R/D, and D/O conditions and also trials of red baseline relations for participants in the latter condition), however, participants learned a series of purely arbitrary conditional discriminations, and had no such straightforward basis for selection available. The virtually errorless maintenance of rhyme baseline relations demonstrated by participants in the R/R, R/O, and R/D conditions during review (phases 2 and 3) did not undermine this interpretation.

Similarly, all participants in the R/R condition met criterion during the first block of extinction full baseline review (Phase 7). Although four participants in the R/O condition also met criterion during that block, only three participants from the R/D condition and the D/O condition performed likewise. All participants met criterion, however, prior to initial emergent testing. Although effects of baseline training order were implicated in the pattern of errors observed in the data of two participants in the R/D condition during this testing block (AF and CP selected only diagonal comparisons on rhyme trials), no significant effects of training order were observed during any subsequent training or testing block.

It seems plausible further to suggest that participants' verbal behaviour was functional during emergent testing (Phase 8). During the first testing block, all participants in the R/R condition but one performed without error, the remaining participant performing likewise during the second block. By contrast, only two participants from the other three conditions (one from the R/O condition and one from the R/D condition) met the 23 of 24 accuracy criterion during initial emergent testing: Only one participant from the D/O condition made fewer than nine errors during either emergent block. Although five of 12 participants in the R/O and R/D conditions performed without error on red (rhyme) trials during the first block of emergent testing, no participant in these conditions performed likewise on blue (non-rhyme) trials. No participant in the D/O condition performed without error on either red (non-rhyme) or blue (non-rhyme) trials during initial emergent testing.

During the first block of emergent testing, no significant effects of nodal distance were observed in the accuracy with which participants in the R/R, R/O, and R/D conditions performed on rhyme trials. In accordance with previous research (Kennedy, 1991; Kennedy et al., 1994), however, participants in the R/O, R/D, and D/O conditions performed with greater accuracy on non-rhyme symmetry trials than on non-rhyme trials of any other relation type. That the latencies of participants in the R/O and R/D conditions were significantly greater on non-rhyme trials of transitivity than on non-rhyme trials of symmetry also supported previous reports of the transitivity latency effect (Dickins et al., 1993).

Although two participants from both the R/O and R/D conditions met criterion for equivalence during tertiary emergent testing (the first emergent block presented on Day Two), no participant in the D/O condition made less than eight errors. One further participant from the R/O condition met criterion for equivalence during final emergent testing, but no participant from either the R/D or D/O conditions performed likewise. By the end of the experiment, therefore, all participants exposed only to rhyme combinations of stimuli had demonstrated equivalence. Only seven of 12 participants exposed to both rhyme and non-rhyme combinations of stimuli, however, had performed likewise. Only one participant exposed only to non-rhyme combinations of stimuli made fewer than five errors during any emergent block.

Initial and secondary generalisation testing (phases 9 and 10) presented novel pictorial stimuli that bore no formal resemblances to each other or to the pictorial stimuli presented previously during the experiment. The names of those stimuli were also phonologically unrelated to the names of any previously presented stimuli. Nevertheless, all participants in the R/R condition selected only comparisons whose names rhymed with those of the samples during initial generalisation testing, regardless of contextual cue. Participants in the D/O condition majoritatively selected non-rhyme comparisons. Almost all participants in the R/O and R/D conditions, however, showed a high level of differential comparison selection: In the presence of the contextual cue that had controlled selection of rhyme stimuli during baseline training, rhyme comparisons were selected, and in the presence of the contextual cue that controlled selection of non-rhyme stimuli during baseline training, non-rhyme comparisons were selected. The results of initial generalisation testing therefore suggest that, because the contextual cues established baseline training had controlled selection of novel stimuli, the generalised contextual control observed had been verbal in nature.

Although the results of secondary generalisation showed greater variability, the functionality of generalised verbal control was again indicated. During blocks 1 and 4, participants in the R/O and R/D conditions majoritatively selected the rhyme key in the response to rhyme compounds and the non-rhyme key in response to non-rhyme compounds. Participants in the R/R condition, however, majoritatively selected the rhyme key in response to both rhyme and non-rhyme compounds during Block 4. Two participants in the D/O condition also selected only rhyme stimuli during that block--again suggesting that when a ready verbal basis for stimulus classification is available (i.e., rhyme), stimuli may be selected on that basis. During blocks 2 and 5, participants in the R/O and R/D conditions majoritatively selected rhyme compounds in response to the rhyme key and non-rhyme compounds in response to the non-rhyme key. Participants in the R/R condition, however, majoritatively selected rhyme compounds in response to both the rhyme and non-rhyme key, and two participants in the D/O condition performed likewise. During blocks 3 and 6, however, the overwhelming majority of participants in the R/R, R/O, and R/D conditions selected rhyme compounds in response to rhyme compounds and non-rhyme compounds in response to non-rhyme compounds. Unexpectedly, however, only two participants in the D/O condition performed likewise.

In summary, therefore, the results of Study Three provide a powerful demonstration of the functionality of participants' verbal behaviour in the formation and generalisation of contextually controlled equivalence classes. Whereas Study Two established contextual control over rhyme and non-rhyme baseline relations, that control was largely superseded by verbal control during emergent testing. The procedural refinements implemented in Study Three, however, resulted in the emergence of equivalence for the majority of participants who were exposed to contextually controlled classes of rhyme and non-rhyme stimuli. Although the principal aim of the experiment was thus achieved, it is not possible to conclude whether the emergence of non-rhyme equivalence classes observed was because of reduction in class sizes, removal of less consistently named stimuli, more extensive exposure to the experimental contingencies, the nature of the contextual stimuli employed, or any interaction of those factors. Although further research will be required to elucidate those potential inter-relationships, two additional findings are, however, indicated: Firstly, participants (even those exposed only to non-rhyme stimulus classes) can demonstrate retention of baseline relations over a minimum period of 16 hr. Secondly, that a number of participants demonstrated equivalence during the first blocks of emergent testing presented on Day Two provides an extension of previous reports of delayed emergence by a demonstration of the delayed emergence of contextually controlled equivalence classes composed of rhyme and non-rhyme stimuli.

7. STUDY FOUR: THINKING ALOUD

7.1 INTRODUCTION

The results of Study Three indicated that participants had named the experimental stimuli without instruction, and that the phonological properties of the names thus given had influenced match-to-sample performance. Once again, these findings were supported by the results of participants' written post-tests. Recent behavioural research (Lane & Critchfield, 1996; Potter et al., 1997; Wulfert et al., 1991, 1994, has suggested, however, that protocol analysis (Ericsson & Simon, 1980, 1993) may provide an additional means by which participants' verbal behaviour during match-to-sample training and testing can be evaluated. Would the collection and analysis of participants' concurrent verbal protocols using a think-aloud procedure provide further insight into the functionality of participants' verbal behaviour indicated by the findings of studies One, Two, and Three?

As Critchfield et al. (1998, p. 435) have observed, "when we study human behavior, we have the luxury of asking our subjects what they know about it", although, as these researchers have acknowledged, "the trustworthiness of such data remains a point of contention"--not least because "there is no hard-and-fast set of rules for doing protocol analysis" (Austin & Delaney, 1998, pp. 42-43). Nevertheless, using a think-aloud procedure, Wulfert et al. (1991, Experiment 2) provided evidence of correlations between participants' verbal behaviour during experimentation and the likelihood of their demonstrating equivalence: Participants who were pre-trained to describe relations between stimuli (e.g., "circle goes with the open triangle") were found to be more likely to demonstrate equivalence than those participants who were pre-trained to describe stimulus compounds (e.g., "together the stimuli look like a house"). Likewise, Wulfert et al. (1994) have reported that protocol analysis can provide insight into the non-emergence of equivalence as a result of failures in stimulus control. Although Lane and Critchfield (1996) have further suggested that the technique may be of use with regard to the investigation of qualitative differences between the component emergent relations of equivalence (cf. Adams et al., 1993a; Fields et al., 1990, 1993), Rehfeldt et al. (1998) have reported that participants' verbal protocols may not necessarily correlate with their experimental performance. The majority of researchers have, however, reported findings to the contrary (e.g., Critchfield & Perone, 1990; Potter et al., 1997; Wulfert et al., 1991, 1994).

Study Four set out to investigate the functionality of participants' verbal behaviour in the contextual control of equivalence classes, by the use of a think-aloud procedure. Although other equivalence studies have assayed participants' verbal behaviour using this technique, no previous research has employed this method to investigate the formation of equivalence classes composed of combinations of stimuli whose names are phonologically related. With regard to this, a mixed experimental design was employed, similar to that reported in Study Three. A number of procedural alterations were implemented, however.

In Study Three, two groups of participants had been exposed to rhyme and non-rhyme combinations of stimuli under the contextual control of the colours of the lines composing those stimuli. Study Four similarly presented all participants with a rhyme combination of stimuli under the contextual control of one such colour, but presented a rearrangement of the same stimuli under the control of the other colour in which the names of the stimuli were alliterative. It was thus proposed to investigate whether a ready verbal basis for stimulus classification other than rhyme would also facilitate the emergence of equivalence. The practical difficulties inherent in devising classes of easily nameable stimuli whose normative names would form the basis of both rhyme and alliteration classes, however, necessitated the employment of two stimuli (*gnat* and *knee*) that had not been presented in the previous study. The stimulus *gnat*, because of its generic nature, was expected to produce variability in the names by which participants referred to it. Although use of this stimulus resulted from purely practical considerations, its presentation was anticipated to provide a potential insight into participants' retrospective renaming of stimuli on the basis of rules derived from the rhyme and alliteration relationships between the names of the other more easily nameable stimuli presented. Another procedural alteration was that trials of baseline relations were presented only in quasi-randomised reinforcement and extinction blocks (cf. Mandell & Sheen, 1994), rather than in the sequential training and testing procedures employed in studies One, Two, and Three.

As Perone (1988, p. 73) has pointed out, an additional use for participants' verbal reports arises "when the reports themselves constitute the object of study". Indeed, as Critchfield et al. (1998, p. 441) have observed, "self-reports can also be viewed as the behavior under investigation". Other researchers have further recommended that experimentation that employs think-aloud procedures should compare the performance of participants that results from those procedures with the performance of participants who are not required to think-aloud (e.g., Hayes, 1986; Hayes et al., 1998; Wulfert et al., 1991). A control condition was therefore proposed, in which participants would be exposed to experimental contingencies identical to those experienced by participants within the think-aloud procedure, but without their being instructed to think-aloud.

Because no equivalence study has addressed the possible effects of instructions to think-aloud in the protocol analysis of emergent behaviour, an additional control condition was employed. Participants in this condition were given instructions to think-aloud, but were only required actually to think-aloud during testing of emergent relations. As during studies One, Two, and Three, written-post-tests were conducted both as a measure of the normativeness of participants' stimulus naming, and as an indicator of potential verbal strategies employed during match-to-sample training and testing. As Critchfield et al. (1998) have observed, converging evidence from different types of self-report may also serve to endorse the validity of those measures.

7.2 METHOD

7.2.1 Participants

Thirty-nine students and staff at the University of Southampton (26 female, 13 male) volunteered to participate in the study and were assigned randomly, but in equal number, to three experimental conditions (protocol, mixed, and no protocol). Aged between 17 and 48 years, all were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 per 30 min, independent of experimental performance.

7.2.2 Apparatus and Setting

Using software specifically designed for equivalence research (Dube & Hiris, 1996), a Power Macintosh® computer presented all stimuli and automatically recorded participants' responses and response latencies. During match-to-sample trials, its 15-in. (38-cm) monitor displayed five transparent keys (4.5 cm key) that were indiscernible against a white background. Sample stimuli were presented on the centre key, and comparisons appeared on three of the four outer keys (see Figure 35). One of the keys, its position varying from trial to trial, always remained blank. Participants were tested individually in an observation laboratory (6 m by 4.5 m) containing a desk on which were placed a sheet of written instructions, a

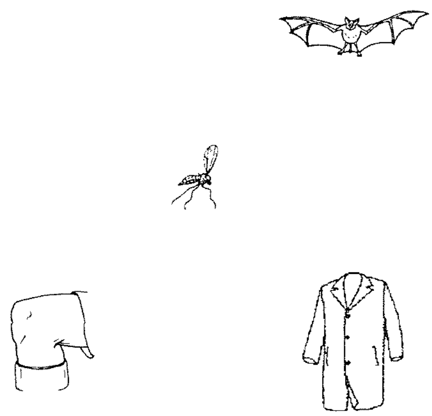

Figure 35. Typical match-to-sample screen display, illustrating an A2B2 trial; *gnat* (sample), *bat* (correct comparison).

computer, monitor, and mouse, and an envelope concealing a pen and post-test booklet for completion subsequent to match-to-sample testing. No keyboard was visible, and responses were made using the mouse. Participants' verbal behaviour was recorded using a ProLine® ADM-1990 boundary microphone and two Sony® EVI-D31 wall-mounted video cameras, all of which were connected to a Panasonic® AG-7350 video recorder and a Panasonic® WJ-MX50 production mixer located in an adjoining room. One camera was focused on the participant from behind the monitor and the other camera, located behind the participant, was focused on the monitor's screen display. The output of the microphone and both cameras was simultaneously mixed and recorded onto a Fuji® Super HG E90 video cassette during each sitting. All participants completed the experiment in one sitting, which never exceeded 1.5 hr duration. If any participant had not satisfied the criteria for baseline training and review within 1 hr, however, the experimenter halted the experiment.

7.2.3 Stimuli and Class Arrangements

Stimuli were nine red and nine blue line drawings of the same easily nameable items (some adapted from Snodgrass & Vanderwart, 1980), the normative names of which were each either three or four letters in length (see Table 15). Two different

Baseline training and emergent testing		
boat	coat	note
gnat	bat	cat
key	knee	bee

Table 15. Normative names of pictorial stimuli used in baseline training and emergent testing.

arrangements of these stimuli (shown in Figure 36) provided the potential for 6 three-member equivalence classes (three rhyme and three alliteration classes). Rhyme classes were composed of stimuli whose names rhymed with each other but were not alliterative, and trials involving these classes (rhyme trials) always presented a sample whose name rhymed with that of the positive comparison but never with those of the two negative comparisons. Alliteration classes were composed of stimuli whose names were alliterative but did not rhyme with each other, and trials involving these classes (alliteration trials) always presented a sample whose name was alliterative with that of the positive comparison, but never with those of the two negative comparisons. Participants in all conditions were exposed to both of these class arrangements, with rhyme and alliteration classes under the contextual control of the red or blue colour of the stimuli, respectively.

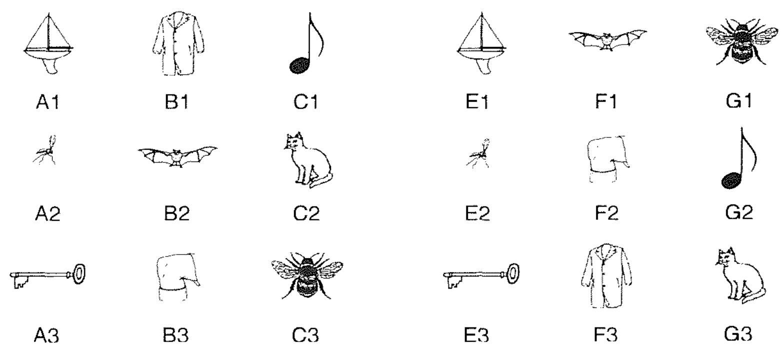

Figure 36. Stimuli and class configurations used in baseline and emergent trials for all conditions. Numbered rows denote classes; lettered columns denote stimuli.

7.2.4 Procedure

7.2.4.1 Verbal instructions.

Prior to experimentation, all participants were informed that their performance would be recorded throughout the experiment by the computer and by the microphone and cameras visible to them. Participants in the no protocol condition were then asked to familiarise themselves with the instructions before them, and left to complete the experiment. Participants in the mixed and protocol conditions, however, were first trained to think-aloud on simple mathematical tasks unrelated to the experiment. To participants in the protocol condition, the experimenter read the following instructions:

> We are interested in understanding how people solve problems, and to help with this, we would like you to think out loud throughout the experiment. So that you understand what I mean by thinking out loud, I will give you an example. Assume I asked you, "How much is 67 plus 28?" Now think out loud so I can hear how you solve this problem.

For participants in the mixed condition, the instructions were the same as those above, except the first sentence was as follows:

> We are interested in understanding how people solve problems, and because of this you may, at some point during the experiment, see a message on the computer screen asking you to think out loud as you continue with experiment.

If any participant stated the solution only ("95"), the experimenter corrected them and modelled an example as follows: "Suppose the problem is 52 plus 49. To solve it, I might think, 52 plus 9 equals 61. Sixty-one plus 40 equals 101. Here's another problem, try thinking out loud again as you solve it". Similar problems were presented until participants thought aloud satisfactorily on two consecutive problems. Participants were then asked to familiarise themselves with the written instructions before them and left to complete the experiment.

7.2.4.2 Written instructions.

Initially, the following text was displayed on the computer's monitor: "When you are familiar with the written instructions, please click on 'Continue' to start the experiment." The written instructions were as follows:

When the experiment begins, and at the start of each subsequent trial, you will see a picture in the middle of the screen in front of you. Use the mouse to click on it. More pictures will now appear in the corners of the screen. Use the mouse to click on one of these. At first, you will receive feedback on your choices, a "beep" for correct, and a "buzz" for incorrect. During later stages of the experiment, you may no longer receive feedback on your choices--the computer will tell you when. Keep on going however, and continue to do the best you can! Please aim to complete the experiment as quickly and accurately as possible. A message on screen will tell you when the experiment is over. When you are ready to start, please click on "Continue". Thank you for participating in this experiment. You are free to leave at any point.

7.2.4.3 Match-to-sample contingencies.

Each trial began with presentation of a sample stimulus, an observing response with the mouse causing comparison stimuli additionally to be displayed. All stimuli remained in view until selection of a comparison caused them immediately to disappear, followed, after a 1-s interval, by presentation of the next trial. Comparison selections made within 0.5 s of presentation had no such consequence, however, and all stimuli remained in view. There was no limit to trial duration. Positions of correct and incorrect comparisons varied pseudo-randomly from trial to trial and, throughout baseline training and emergent testing, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3). At no point during the experiment did the location of the correct comparison remain constant for more than two consecutive trials, or the same sample stimulus appear for more than two trials consecutively. No more than two rhyme or alliteration trials were ever presented consecutively.

All participants were exposed to the same performance-contingent training and testing programme. During reinforcement training, selection of class-consistent comparison stimuli was followed by a beep and the word "CORRECT" displayed on the screen. Other choices resulted in a buzz and a darkened screen. During testing without reinforcement, however, the only consequence of a response was the presentation of the next trial. The overall procedure for all conditions was designed as a series of successive training and testing blocks, the details of which are given below.

7.2.4.4 Phase 1: Train rhyme (AB, BC) and alliteration (EF, FG) baseline relations.

Initially, all AB, BC, EF, and FG relations were trained with reinforcement in 12-trial blocks, with each of the six rhyme and six alliteration baseline relations presented in pseudo-random order once every block. On completion of one reinforcement trial block with 100% accuracy, all rhyme and alliteration baseline relations were reviewed in extinction (Phase 2). Table 16 shows all baseline trial configurations used during baseline training and review.

Trained Relations											
AB			BC			EF			FG		
Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
A1	B1	B2, B3	B1	C1	C2, C3	E1	F1	F2, F3	F1	G1	G2, G3
A2	B2	B1, B3	B2	C2	C1, C3	E2	F2	F1, F3	F2	G2	G1, G3
A3	B3	B1, B2	B3	C3	C1, C2	E3	F3	F1, F2	F3	G3	G1, G2

Table 16. Baseline trial configurations, using single-sample and four comparison displays.

7.2.4.5 Phase 2: Review rhyme and alliteration baseline relations.

Otherwise identical to Phase 1, all trials during this phase were completed in the absence of reinforcement. If performance remained at 100% accuracy over the first 12-trial block, emergent testing commenced (Phase 3). If criterion was not met in this block, however, baseline relations were retrained with feedback (Phase 1). On completion of one block of reinforcement trials to the same criterion, maintenance of the baseline relations was again assessed in extinction (Phase 2). Retraining of baseline relations and review of their maintenance continued in this way until criterion was met in extinction. When participants in the mixed condition met criterion for baseline relations in extinction, the following message was displayed on-screen prior to commencement of emergent testing.

Please now THINK OUT LOUD as you continue with the experiment!

7.2.4.6 Phase 3: Emergent testing.

	Emergent relations								
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
Symmetry									
BA	B1	A1	A2, A3	B2	A2	A1, A3	B3	A3	A1, A2
CB	C1	B1	B2, B3	C2	B2	B1, B3	C3	B3	B1, B2
FE	E1	E1	E2, E3	F2	E2	E1, E3	F3	E3	E1, E2
GE	G1	F1	F2, F3	G2	F2	E1, F3	G3	F3	F1, F2
Transitivity									
AC	A1	C1	C2, C3	A2	C2	C1, C3	A3	C3	C1, C2
EG	E1	G1	G2, G3	E2	G2	G1, G3	E3	G3	G1, G2
Equivalence									
CA	C1	A1	A2, A3	C2	A2	A1, A3	C3	A3	A1, A2
GE	G1	E1	E2, E3	G2	E2	E1, E3	G3	E3	E1, E2

Table 17. Emergent trial configurations, using single-sample and three comparison displays.

All possible emergent relations except reflexivity were presented, in pseudo-random order, in a maximum of four 24-trial blocks (see Table 17). Generalised identity-matching repertoires were assumed. If participants satisfied the criterion of a minimum of 23 of 24 class-consistent responses in any one emergent testing block, match-to-sample testing ended. An on-screen message informed participants of this and asked them to complete the post-test in the envelope before them. If, however, criterion had not been achieved at the end of the first two consecutive blocks of emergent testing, baseline relations were again reviewed, without feedback (Phase 2). If criterion had not been achieved at the end of one test trial block, baseline relations were again retrained (Phase 1). Review and retraining of baseline continued in this way until all relations were again demonstrated with 100% accuracy over one block of test trials. The second two consecutive blocks of emergent testing were then presented. Consequent to completion of the fourth testing block, match-to-sample testing ended and an on-screen message informed participants of this, regardless of performance.

7.2.4.7 Phase 4: Naming post-test.

Subsequent to completion of emergent testing, all participants completed a written post-test that was designed to elucidate their verbal behaviour during the experiment. Although otherwise identical for participants in all conditions, the post-test booklet completed by participants in the mixed and protocol conditions contained two extra questions (see below). All booklets were headed by the following instructions:

> Printed below are the pictures that you saw during the experiment. Did you name any of them, mentally or out loud, or refer to them in any way during testing? If you did, please write beside each picture the name you used for it. If you referred to any picture by more than one name during the experiment, please indicate all the names you used, in the order that you used them. If you did not refer to a picture in any such way, please leave the space beside it blank. Please note: If the names you used changed during the course of the experiment, you will have opportunity to indicate why they changed later in the post-test.

The experimental stimuli used during baseline and emergent testing were then presented, each followed by a blank space and dotted line. Beneath these, the following questions were presented, each followed by a blank space:

> Did you use any strategies, 'tricks', or rules to learn and remember which of the above pictures went together? If you did, please enter a brief description below. If the strategies you used differed depending on whether the pictures were RED or BLUE, please describe how they differed. If you did not use any such strategies, please leave the space below blank and complete the rest of the post-test.

and

> Did any of the strategies, 'tricks', or rules that you may have used, or any of the names by which you may have referred to pictures, change during the course of the experiment? If they did, please enter a brief description below.

For participants in the protocol and mixed conditions, the following questions were presented next:

You were asked to think out loud during the experiment. Do you feel that doing so affected your performance in any way? If you think it did, please indicate how.

and

Do you feel that what you said was an accurate reflection of what you were doing and thinking during the experiment? If not, how did it differ?

For participants in all conditions the final question was as follows:

If there are any other observations you would like to make regarding your performance during the experiment, please note these below.

7.3 RESULTS

Although all participants in the no protocol condition met the criteria for baseline training and maintenance and demonstrated equivalence, only eight of 13 participants in the protocol condition ever met the former criterion, and only four of those demonstrated equivalence. All participants in the mixed condition met the baseline criteria, but only nine of those ever met criterion for equivalence. Individual participants' trials and errors during all phases of match-to-sample training and testing are presented in Appendix G. Session durations for all participants are presented in Appendix H.

7.3.1 Phase 1: Train Rhyme and Alliteration Baseline Relations

Acquisition of baseline relations was easiest for participants in the no protocol condition and was most difficult for those in the protocol condition. No participant in the no protocol condition required more than 18 12-trial blocks to meet criterion for baseline relations with reinforcement, and the smallest number of blocks required was two. Although the smallest number of blocks required to meet the same criterion in both the protocol and mixed conditions was three, the greatest number of blocks required in the latter condition was 31. Of the seven participants in the protocol condition who ever met criterion, the greatest number of blocks required was 25. The greatest number of blocks completed by any of the other five participants before termination of the experiment was 47. The participation of three participants in the protocol condition (KS, LD, and RC) was terminated at 1 hr. For three others (LC, OH, and SL), the experiment was terminated on request at 50 min, 46 min, and 34 min, respectively.

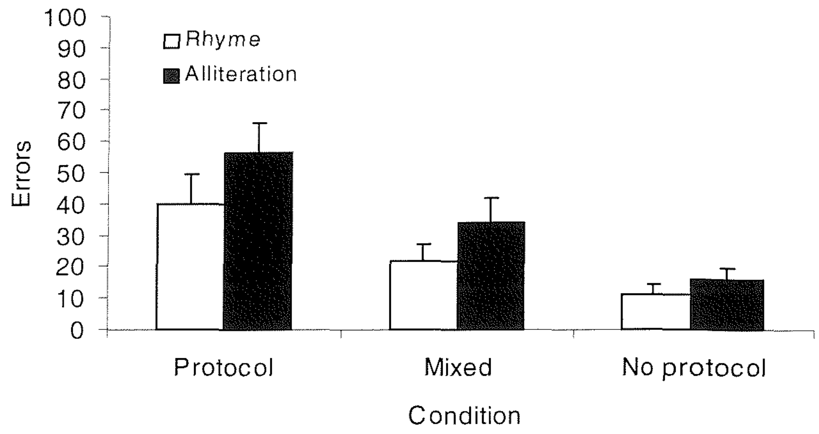

Figure 37. Mean trials and errors (+*SE*) for all participants in protocol, mixed, and no protocol conditions initially to meet criterion for baseline training (Phase 1) or until termination of the experiment.

Participants in the no protocol condition required a mean of 94.2 trials (*SD* = 61.9) initially to meet criterion for baseline relations with reinforcement, whereas those in the mixed condition required a mean of 167.1 trials (*SD* = 106.3). Participants in the protocol condition were exposed to a mean of 230.8 trials (*SD* = 158.3) prior to meeting the same criterion or termination of the experiment. These data were subjected to a between-participant analysis of variance, in which the factor was condition (protocol, mixed, no protocol). A significant main effect of condition was observed, $F(2) = 4.54, p < .05$, Scheffé tests (at an alpha level of .05) further indicating that although significantly fewer trials were required in the no protocol condition than in the protocol condition initially to meet criterion, the number of trials required in those conditions did not differ significantly from those required in the mixed condition. Figure 37 shows the mean number of errors made on rhyme and alliteration trials by all participants during baseline training. Participants' error scores were subjected to a mixed design analysis of variance in which the between-participant factor was condition (protocol, mixed, no protocol) and the within-participant factor was context (rhyme, alliteration). Significant main effects of condition, $F(2, 36) = 6.62, p < .05$, and context, $F(1, 36) = 27.75, p < .001$, were observed, and Scheffé tests (at an alpha level of .05) again indicated that although participants in the no protocol condition made significantly fewer errors on rhyme and alliteration trials than participants in the protocol condition, the error scores of participants in those conditions did not differ significantly from those of participants in the mixed condition.

7.3.2 Phases 2 and 3: Review Baseline Relations and Emergent Testing

7.3.2.1 Review prior to initial emergent testing.

All participants in the no protocol condition but three met the errorless criterion for maintenance of rhyme and alliteration baseline relations in extinction in the first 12-trial review block and proceeded immediately to emergent testing (Phase 3). All remaining participants in that condition met criterion in their second review block, and only one of those participants required more than a single additional reinforcement block to re-establish baseline relations to criterion. In the mixed condition, eight participants met criterion for maintenance of baseline relations in extinction in the first review block, and the five others met criterion during their second extinction review block. Four of these participants required two additional reinforcement blocks or more to re-establish baseline relations to criterion. Of the eight participants in the protocol condition who ever met criterion for baseline relations with reinforcement, five met criterion for baseline relations in extinction in the first 12-trial review block. Two others met criterion in the second block, and neither of these participants required more than one reinforcement training block to meet criterion in extinction. The experiment was terminated during the other participant's first block of extinction baseline review.

7.3.2.2 Initial emergent testing (blocks 1 and 2).

During the first block of emergent testing, 10 of 13 no protocol participants fulfilled the 23 of 24 criterion for equivalence; eight performed errorlessly. The remaining three participants performed without error during the second testing block. By contrast, only five participants from the mixed condition achieved criterion during the first block of testing, although another two did so in the second block. In the protocol condition, four of the seven participants who had met criterion for baseline relations in extinction met criterion during initial emergent testing, all in the first block. The top two sections of Figure 38 show means of percentage error scores on rhyme and alliteration trials for every participant in all conditions who was exposed to initial emergent testing.

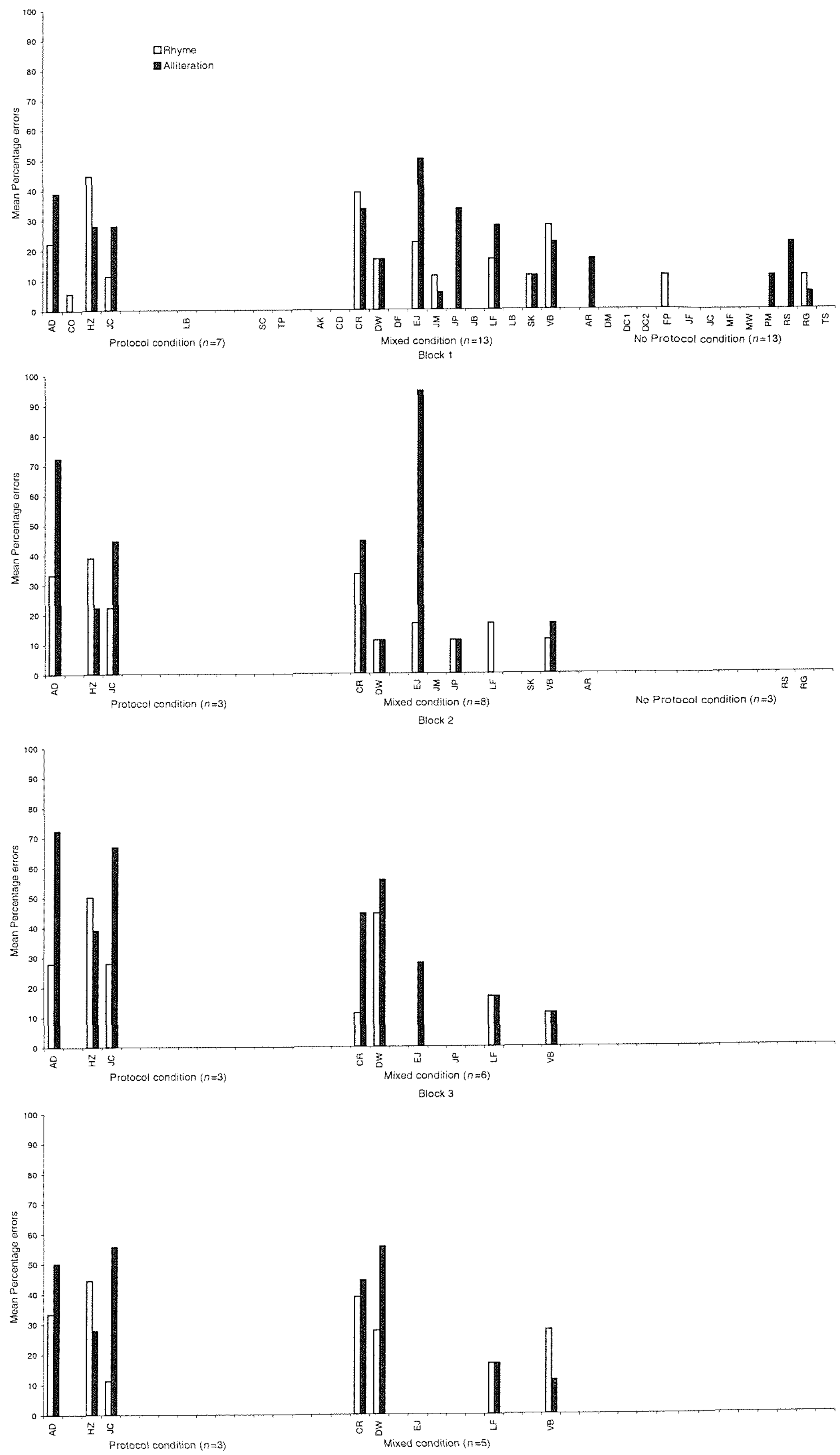

Figure 38. Means of percentage error scores of all participants in the protocol, mixed, and no protocol conditions during initial and final emergent testing (Phase 3). Absence of initials indicates that trials were not presented to that participant.

7.3.2.3 Review prior to final emergent testing.

Because all participants in the no protocol condition had demonstrated the formation of contextually controlled equivalence classes during initial emergent testing, none received further baseline review. Of the six participants in the mixed condition who had failed to meet criterion during initial emergent testing, only one demonstrated continued baseline maintenance in extinction, although four of the remaining participants met criterion for baseline relations in extinction in their second review block. The other participant met criterion in the third review block. Five of these participants required two additional blocks of reinforcement baseline training or more to meet criterion. Of the three participants in the protocol condition who had failed to meet criterion during initial emergent testing, one demonstrated continued baseline maintenance in the first extinction review block. Another met criterion in the second block, and the other did so in the third. Neither of the latter participants required less than four reinforcement training blocks to meet criterion.

7.3.2.4 Final emergent testing (blocks 3 and 4).

None of the participants in the protocol condition who had failed to meet criterion during initial emergent testing met criterion during final emergent testing. Of the six participants in the mixed condition who had failed to meet the same criterion, only two met criterion during final emergent testing, one in the first block and one in the second. The bottom two sections of Figure 38 show means of percentage error scores on rhyme and alliteration trials for every participant in the mixed and protocol conditions who was exposed to final emergent testing.

7.3.3 Phase 4: Naming Post-test

Because three participants in the protocol condition (KS, LD, and RC) had failed to demonstrate mastery of baseline relations within 1 hr, their participation in the experiment was terminated. For three other participants in that condition (LC, OH, and SL), the experiment was terminated on request prior to mastery of baseline relations. Owing to this, none of these participants completed written post-tests.

All other participants' post-tests suggested a high degree of normative stimulus naming: All participants in the no protocol condition, and all participants in the mixed condition but two, reported using only normative stimulus names during the experiment. Ten participants in the no protocol condition and seven in the mixed condition, however, indicated that they had initially referred to *gnat* by one or more of a variety of non-normative names (i.e., *ant, flea*, *bug*, *mosquito*, or *insect*). Three participants in the former condition and one in the latter condition further indicated that they had initially referred to *boat* by a non-normative name (i.e., *yacht* or *ship*), and one participant in the no protocol condition further reported that he had initially referred to *coat* as *tuxedo*. Two participants in the mixed condition additionally indicated that they had never referred to *gnat* by its normative name. Only one participant in the protocol condition (SC) indicated that he had referred to all stimuli by their normative names throughout the experiment. Two other participants in the protocol condition (HZ and TP) noted that they had initially referred to *boat* by a non-normative name, and three (HZ, JC, and TP) indicated that they had initially referred to *coat* as either *shirt* or *jacket*. One participant in the protocol condition (HZ) reported that he had referred to *note* as "*musical symbol that rhymes with coat*" throughout the experiment, and five participants (AD, CO, HZ, JC, and LB) indicated that they had initially referred to *gnat* by a non-normative name.

All 13 participants in the no protocol condition indicated that they had selected stimuli on the basis of rhyme during rhyme trials, and on the basis of alliteration during alliteration trials. Ten of these participants additionally indicated that they had retrospectively changed the name by which they had initially referred to *gnat* on the basis of the rhyme and alliteration of that stimulus' normative name with the normative names of the other stimuli presented. Nine participants in the mixed condition indicated that they had selected stimuli on the basis of rhyme and alliteration. Three other participants in that condition, however, reported that although they had selected stimuli on the basis of rhyme on rhyme trials, they had majoritatively selected stimuli on the basis of a shared first letter on alliteration trials. One other participant in the mixed condition reported using intraverbal strategies (e.g., *mosquito bites knee*) to select stimuli on alliteration trials.

All participants in the mixed condition reported that their verbalisations had been an accurate reflection of what they had done and thought during emergent testing, and five also reported that thinking aloud had not affected their performance. Two others, however, reported that thinking aloud had slowed down their performance. Three further participants indicated that thinking aloud had been beneficial to their performance, and two others noted that although thinking aloud had initially impaired their performance, it had subsequently been beneficial.

Although five participants in the protocol condition (CO, HZ, LB, SC, and TP) indicated that they had selected comparisons differentially on the basis of rhyme and alliteration, only two of these participants (CO and LB) reported that the non-normative name by which they had initially referred to *gnat* had changed on the basis of the rhyme and alliteration between that stimulus' normative name and the normative names of the other stimuli in the experiment. HZ additionally noted selecting stimuli on alliteration trials on the basis of intraverbal strategies when there was "no match" between the names of stimuli, and two other participants in this condition (AD and JC) noted selecting stimuli on the basis of intraverbal strategies alone on alliteration trials.

Six participants in the protocol condition (AD, CO, HZ, JC, LB, and TP) noted that thinking out loud had been beneficial to their performance, although two of these participants (AD and TP) also indicated that it had initially been detrimental. The former six participants also reported that their verbalisations had been an accurate, or "fairly accurate" reflection of what they had thought and done during the experiment. One participant from the no protocol condition and another from the mixed condition also observed that they had found it easier to select stimuli on the basis of rhyme than on the basis of alliteration.

7.3.4 Verbal Protocols

Participants' verbalisations were transcribed directly from videotape on a trial by trial basis using a variant of the observational recording technique (Reese, Howard, & Reese, 1978), and then summarised. Because the transcription and analysis of participants' verbal behaviour are highly labour intensive, the protocols of six participants in the protocol condition only were transcribed and analysed (cf. Wulfert et al., 1991). Two of these participants (SC and TP) had completed the experiment and demonstrated equivalence. Two others (JC and AD) had completed the experiment but had not demonstrated equivalence. The final two participants (LC and RC) had never met criterion for baseline relations.

Participant SC

Participant SC required only 36 trials to meet criterion for baseline training and demonstrated equivalence during his first block of emergent testing. SC never non-normatively named a stimulus. During his second rhyme trial, he stated "obviously rhyming", and selected only rhyme comparisons on the following two trials regardless of contextual cue. On his 17th trial (alliteration), SC stated "not that one, it rhymes" prior to selecting a non-rhyme stimulus. During his 23rd trial (rhyme), SC stated that "there is some colour thing going on", and on his next alliteration trial said "it's spelling, pronunciation". SC made no further errors during the experiment. During each of his last eight baseline training trials and first eleven subsequent baseline review trials, SC named both the sample and comparison stimuli prefixed by the colour of the contextual cue (e.g., "blue cat, blue key"). Thereafter, SC stated only the names of the sample and comparison stimuli. The session duration for this participant was 17 min.

Participant TP

Participant TP required 84 trials to meet criterion for baseline training but performed errorlessly throughout the rest of the experiment, except for a single error during his first block of baseline review. TP thought aloud infrequently, but during his second baseline training trial (alliteration) indicated selecting a comparison on the basis of spatial position (i.e., "top right"). On his fifth trial (rhyme), he also indicated physical similarity between stimuli (i.e., "have wings"). TP's verbalisation on his 22nd trial (alliteration) again suggested comparison selection on the basis of a spatio-temporal strategy (i.e., "it's position on the screen). During his 55th trial (alliteration), however, TP said "blue is phonetic start of word sound", and on his 58th trial (rhyme) stated "red is functionality, no--rhyme!". From his 61st to his 65th baseline training trial TP named both sample and comparison stimuli out loud (e.g., "knee, bee"), and continued to do so until his 70th trial. TP did not think-aloud again during the experiment. The session duration for this participant was 28 min.

Participant JC

Participant JC completed the experiment, but did not make fewer than four errors during any emergent testing block. In total, she was exposed to 468 trials. JC referred to *gnat* as *fly* throughout the experiment, and referred to *coat* as *jacket* until her 48th baseline training trial. Subsequent to this, JC named *coat* normatively. During her first trial, and on another 108 of her first 150 reinforcement training trials, JC named both the sample and the comparison that she selected without reference to the colour of the contextual cue. On another 20 of those trials, JC named only the sample stimulus. During the other 21 trials, JC either did not think-aloud, or stated "can't see any connection", or "don't know why those go together". On each of her final 318 trials, JC always named the sample and the comparison that she selected, prefixed by the colour of the contextual cue (e.g., "red boat, red coat", or "red boat goes with red coat"). The session duration for this participant was 1 hr.

Participant AD

Participant AD completed the experiment, but did not make fewer than six errors during any emergent testing block. In total, AD was exposed to 492 trials. Initially, AD referred to *coat* as *overcoat*, *boat* as *sailing boat*, and *note* as *musical note*. Subsequent to his 28th baseline training trial, however, he named all three stimuli normatively. AD referred to *gnat* as *mosquito*, however, throughout the experiment. From his first trial, AD named both the sample stimulus and the comparison that he selected, prefixed by the colour of contextual cue (e.g., "red bat, red cat"). During his first 12 trials only, these statements were linked by "I'm going to click on" (e.g., "red cat--I'm going to click on red bat"). The session duration for this participant was 70 min.

Participant LC

Participant LC was exposed to 203 reinforcement trials, but never met criterion for baseline training. LC thought aloud during his first 102 trials, and named every stimulus normatively except *gnat*, which he always referred to as *fly*. During his final 101 trials, LC did not think-aloud. LC selected a rhyme comparison on his first (rhyme) trial, and stated "coat rhymes with note" on the second (alliteration) trial, prior to selecting the rhyme comparison. On his second alliteration trial, LC stated "bat and boat begin with 'B'", and selected the alliteration comparison. LC stated "begins with 'B'" on the next (rhyme) trial, and again selected the alliteration comparison. During the remaining 97 trials on which LC thought aloud, his verbalisations took three main forms: Naming of the sample stimulus, or the sample and one or more of the comparison stimuli (38 trials). Statement of the colour of the contextual cue only (27 trials). A statement of the on-screen position of the comparison that he selected (e.g., "top left") prefixed, on 21 trials, by the name of the sample stimulus (e.g., "boat, top left"). During the other 11 trials, LC either stated "begins with 'B'", or made verbalisations such as "no idea", or "don't understand". The session was terminated on the request of this participant at 50 min.

Participant RC

Participant RC completed 569 reinforcement trials, but never met criterion for baseline relations. RC did not think-aloud during his last 195 trials or on 26 other trials during the experiment. He named four stimuli non-normatively on all the trials during which he thought aloud, referring to *coat* as *overcoat*, *note* as *music*, and *gnat* as *fly*. During his first 18 trials, he also referred to *boat* as *toy yacht*, but always referred to it thereafter as *yacht*. On his seventh trial, RC stated "bats and bees fly, so fly together". During 24 of the 348 trials on which RC thought aloud, he named only the sample stimulus. On 309 trials, RC named both the sample and the comparison (e.g., "bat to cat", or "cat goes with bat"). On 101 of those trials his statements were prefixed by the colour of the contextual cue (e.g., "red bat to red cat", or "red cat, red bat"). On the other 14 trials, he either stated "wrong" subsequent to selecting an incorrect comparison, or "don't know why those go together" subsequent to selecting a correct comparison. This participants' session was terminated by the experimenter at 1 hr.

7.3.5 Session Duration

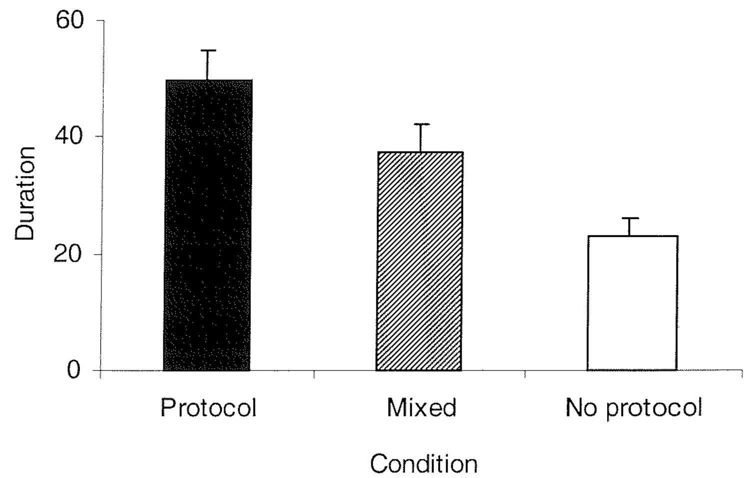

Figure 39. Mean session duration in minutes (+*SE*) in all conditions until completion or termination of the experiment.

No participant in the no protocol condition required longer than 45 min to complete the experiment, and the shortest session duration in this condition was 8 min (DC2). By contrast, the greatest amount of time required to complete the experiment by a participant in the protocol condition was 88 min (HZ), and the shortest session duration in this condition was 17 min (HZ). In the mixed condition, the shortest duration was 13 min (LB) and the greatest duration was 1 hr (EJ). Session durations were rounded to the nearest minute for all participants. Figure 39 shows mean session durations for participants in all conditions to completion or termination of the experiment.

7.4 DISCUSSION

The results suggested that participants named the experimental stimuli without instruction so to do, and that the phonological properties of the names thus given had influenced match-to-sample performance. All participants were exposed to the same match-to-sample procedure during training of rhyme and alliteration baseline relations (Phase 1), and all participants in the mixed and no protocol conditions met criterion during that phase. Only eight of 13 participants in the protocol condition, however, performed likewise. Although the trial and error scores of participants in the mixed and no protocol conditions did not differ significantly from each other during baseline training, instructional effects were nevertheless suggested by the greater number of trials required and errors made by participants in the mixed condition. That the trial and error scores of participants in the protocol condition were significantly greater than those of participants in the no protocol condition may further indicate that both instructions to think-aloud, and thinking aloud itself, had been detrimental to the performance of participants in the former condition.

The results of baseline review (Phase 2) and emergent testing (Phase 3) supported this conclusion: All participants in the no protocol condition met the 23 of 24 accuracy criterion without error during initial emergent testing (blocks 1 and 2). Only seven participants in the mixed condition performed likewise, however, although two further participants in that condition performed without error during final emergent testing (blocks 3 and 4). Only four of seven participants in the protocol condition who had met criterion for baseline review ever met criterion for equivalence. By the end of the experiment, therefore, all 13 participants in the no protocol condition had demonstrated the formation of contextually controlled equivalence classes. Only nine participants in the mixed condition, however, and four participants in the protocol condition, had performed likewise.

These findings were unexpected, because a large amount of research has previously indicated that the implementation of think-aloud procedures does not affect participants' experimental performance (Ericsson & Simon, 1980, 1993, 1998; Hayes, 1986; Hayes et al., 1998). Russo, Johnson, and Stephens (1989) have pointed out that such evidence may not be conclusive, however, and have themselves reported the reactivity of thinking aloud during problem-solving tasks. Short, Evans, Friebert, and Schatschneider (1991) have further demonstrated that thinking aloud can facilitate performance during such tasks. The results of Study Four indicate that the implementation of a think-aloud procedure can also reduce the accuracy with which participants perform on match-to-sample tasks during equivalence training and testing. Various researchers (e.g., Payne, Braunstein, & Carroll, 1978; Ransdell, 1995) have additionally noted that thinking aloud can slow down participants' performance, and the session durations of participants in the protocol, mixed, and no protocol conditions observed during Study Four supported this conclusion.

Although Ericsson and Simon (1984, 1993) have indicated quantitative techniques by which participants' verbal protocols can be coded subsequent to transcription, participants verbalisations during Study Four were not so coded for a number of reasons: Firstly, the use of such techniques is heavily dependent upon *a priori* theoretical assumptions regarding the data observed (cf. Critchfield et al., 1998; Wulfert et al., 1991), and such assumptions necessarily influence conclusions generated from the outcomes of that coding. Secondly, purely quantitative analyses may obscure changes in participants' performance over time, and thus also obscure important features of the data. This consideration was of particular relevance during Study Four because sudden changes in verbal behaviour were observed during the course of the experiment. Thirdly, a more descriptive approach to the data served to reduce potential confounds introduced by variations in the frequency with which participants thought aloud during the experiment. Comparison of the verbalisations of four participants in the protocol condition with their written post-tests nonetheless indicated strong correlations between their observed and reported verbal behaviour. Both measures were also strongly correlated with participants' match-to-sample performance.

Both measures additionally suggested potential explanations for the differences in accuracy with which participants in the protocol condition performed, and also supported previous reports of the contraction of verbalisations during the course of experimentation (Horne & Lowe, 1996; Wulfert et al., 1991, 1994). Participant SC's verbalisations indicated that he had named stimuli normatively throughout the experiment, and also suggested that from early in the experiment his behaviour had been rule-governed: SC stated "obviously rhyming" immediately prior to consistent selection of rhyme stimuli and "it's spelling, pronunciation", prior to consistent selection of alliteration stimuli. The statement "there's some colour thing going on" was also correlated with consistent contextually controlled selection of rhyme and alliteration stimuli. Conversely, participant TP's initially low accuracy performance was correlated with verbalisations that suggested comparison selection on the basis of physical similarity or spatio-temporal location of stimuli. This participant also made statements suggestive of the development of rhyme and alliteration rules, however, immediately prior to consistent differential selection of rhyme and alliteration stimuli.

The verbalisations of participants JC and AD were not suggestive of the use of such rules, however, and these participants also did not meet criterion for equivalence. Although participant LC's verbalisations initially suggested comparison selection on the basis of rhyme, this participant never met criterion for baseline relations. On his first rhyme trial, LC selected a rhyme stimulus, and on the next alliteration trial stated "coat rhymes with note" prior to selecting the rhyme comparison. On his second alliteration trial he stated "bat and boat begin with 'B'", and on the first subsequent rhyme trial stated "begins with 'B'" prior to selecting a non-rhyme comparison. On all subsequent trials during which he thought aloud, this participant did not state either rhyme or alliteration rules. Participant RC also never met criterion for baseline relations, and performed with the lowest level of accuracy of any participant. He was also observed to have non-normatively named more stimuli than any other participant. Although one of RC's statements at the beginning of the experiment suggested selection of stimuli on the basis of physical similarity, none of his other statements during the experiment indicated comparison selection on the basis of any verbal rules.

The verbalisations of two participants who demonstrated equivalence therefore indicated that they had named all stimuli normatively and that their performance was rule-governed from early in the experiment. The verbalisations of four participants who did not demonstrate equivalence indicated that they had consistently non-normatively named one or more stimuli, and that they had not made statements suggestive of the consistent use of verbal rules. It may therefore be suggested that when even a small number of stimuli were non-normatively named by participants throughout the experiment, use of those names resulted in the non-emergence of verbal rules among those participants.

In summary, therefore, the results of Study Four provide a powerful demonstration of the functionality of participant's verbal behaviour in the formation of contextually controlled equivalence classes composed of stimuli whose names are phonologically related. The comparison of participants' concurrent protocols with their written post-tests provides converging evidence of the validity of verbal reports as a measure of participants' verbal behaviour during experimentation. Nevertheless, the results of match-to-sample training and testing suggest that the implementation of such a procedure can affect participants' performance in two ways; through the task demands of the procedure itself, and through effects of the instructions inherent in its use. Further research will be required, however, to elucidate the potential relationships between those two factors.

8. STUDY FIVE: EQUIVALENCE AND REINFORCEMENT TRAINING

8.1 EXPERIMENT ONE: INTRODUCTION

The results of Study Four again suggested that participants had named the experimental stimuli without instruction so to do, and that the phonological properties of the names thus given had influenced match-to-sample performance. These findings were again supported by participants' written post-tests, and also by observation of participants' verbal behaviour during experimentation. Both measures additionally suggested that the formation of contextually controlled equivalence classes can be disrupted both by the instructional effects and the task demands inherent in the use of a think-aloud procedure.

The principal aim of studies One to Four was to investigate the potentially functional role of participants' verbal behaviour in the formation and generalisation of equivalence classes. Because a fundamental aspect of the definition of such classes is that they are composed of stimuli that have become equivalent as a result of specific patterns of baseline training (Saunders & Green, 1992; Sidman, 1994, 1997a; Sidman & Tailby, 1982), the definitional relations of reflexivity, symmetry, and transitivity necessarily describe the reflexive, symmetric, and transitive properties of relations trained between stimuli. Various researchers have, however, suggested that there may be qualitative differences between the emergent relations of equivalence (e.g., Fields et al., 1990; Fields & Verhave, 1987; Saunders & Green, 1992), and such suggestions have been supported by analysis of the structure of equivalence classes with regard to the effects of nodal distance (e.g., Bentall et al., 1993, 1999; Dickins et al., 1993; Fields et al., 1993, 1984; Kennedy, 1991; Kennedy et al., 1994).

As noted previously, equivalence classes can be defined as arbitrary classes, in that they are composed of finite numbers of physically unrelated stimuli. By contrast, feature classes can be defined as classes composed of potentially unlimited numbers of stimuli whose class membership is dependent upon physical similarity (Stromer & Mackay, 1996). A number of studies have indicated that equivalence classes can generalise to novel stimuli on the basis of physical similarity between the latter stimuli and the stimuli already participating in equivalence classes (e.g., Adams et al., 1993b; Fields et al., 1996, 1997, 1991; Meehan & Fields, 1995).

Other research has indicated that equivalence classes can form in the absence of differential consequences. Saunders et al. (1988), for instance, employed intellectually-disabled participants with prior experience of equivalence training and testing to demonstrate the merger of equivalence classes composed of arbitrary visual forms in the absence of direct reinforcement (i.e., reinforcement consequent to conditional selection of those specific stimuli). When stimuli that already participated in equivalence classes were presented as samples, and stimuli that already participated in other equivalence classes were presented as comparisons on every trial using a two choice match-to-sample procedure, equivalence was observed to have emerged between stimuli in the former and latter classes on the basis of the consistent, yet unreinforced, conditional discriminations to which participants had been exposed.

Harrison and Green (1990) extended these findings by demonstrating that mainstream adults and children with no prior experimental history of reinforced conditional discrimination training can also demonstrate equivalence between arbitrary visual forms in the absence of direct reinforcement. To achieve this, a two-choice match-to-sample procedure was employed that presented stable sample-correct comparison pairings in the presence of a constantly changing incorrect comparison. As Sidman (1992, pp. 24-25) has pointed out, "it remains now to determine how far this can be carried... All that should be required [to generate bidirectional conditional relations between stimuli] is to ensure that selections on the basis of any undesired relations are possible on only some of the trials, and that selection on the basis of the desired relations is possible on every trial."

Might there be another way to produce equivalence in the absence of direct reinforcement, however, using verbally able, yet experimentally naïve, participants? Might such participants' verbal behaviour alone prove sufficient to produce equivalence between formally unrelated visual stimuli whose names are phonologically related? Such a finding would provide further evidence of the functionality of participants' verbal behaviour during equivalence research, and might additionally question the necessity of reinforcement baseline training as a definitional requirement of stimulus equivalence. Such a finding might also indicate that because of participants' verbal behaviour, stimuli can participate in feature classes on the basis of phonological, as well as formal, similarity.

No published research has attempted to investigate the formation of verbally controlled bidirectional stimulus classes in the absence of direct reinforcement. Experiment One was proposed so to do, initially by exposing verbally able adult participants to reinforcement sample-correct comparison pairings involving easily nameable, yet formally unrelated, pictorial stimuli whose names rhymed with each other. In order to assess the parameters of reinforcement required to generate such bidirectional relations, participants were exposed to different numbers of reinforcement training trials prior to testing in extinction. Subsequent test trials presented arrangements of other easily nameable, yet formally unrelated, pictorial stimuli as samples and comparisons in the absence of reinforcement. On every trial, however, the normative name of one comparison stimulus rhymed with that of the sample. The stimuli presented during Study One were employed throughout extinction testing, and class arrangements were identical to those to which participants in the rhyme condition during Study One were exposed. As a measure of the normativeness of participants' stimulus naming, and as an indicator of potential verbal strategies employed by participants during match-to-sample training and testing, written post-experimental tests of naming were carried out.

8.2 METHOD

8.2.1 Participants

Three female students at the University of Southampton (JM, KS, and CB) volunteered to participate in the study. Aged between 19 and 21 years, all were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 per 30 min, independent of experimental performance. Data from two additional participants (CY and KY), whose first language was not English, are considered separately.

8.2.2 Apparatus and Setting

Using software specifically designed for equivalence research (Dube & Hiris, 1996), a Power Macintosh® computer presented all stimuli and automatically recorded participants' responses and response latencies. During emergent testing, its 15-in. (38-cm) monitor displayed five transparent keys (4.5 cm square) that were indiscernible against a white background. Sample stimuli were presented on the centre key, and comparisons appeared on the four outer keys (see Figure 40). During generalisation testing, one of the outer keys, its position varying from trial to trial, always remained blank. During reinforcement training, three of the outer keys, their positions varying from trial to trial, always remained blank. Participants were tested individually in a small windowless cubicle (1.5 m by 2.9 m) containing a desk, on which were placed a sheet of written instructions, the computer, monitor, and mouse, and an envelope concealing a pen and post-test booklet for completion subsequent to match-to-sample testing. No keyboard was visible, and responses were made using the mouse. All participants completed the experiment in one sitting, which never exceeded 1 hr duration.

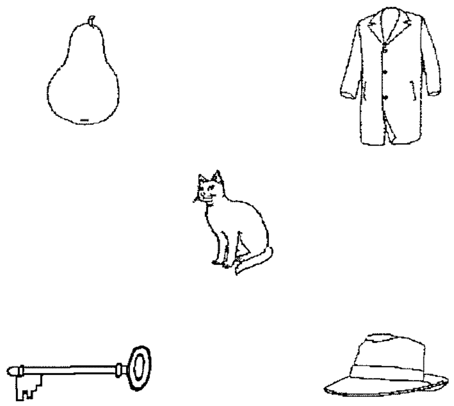

Figure 40. Typical match-to-sample screen display, illustrating a C3D3 trial; *cat* (sample), *hat* (correct comparison).

8.2.3 Stimuli and Class Arrangements

Initial and secondary emergent testing				Reinforcement training		Generalisation testing		
boat	goat	note	coat	sock	lock	can	man	fan
flea	tree	bee	key	car	star	dog	frog	log
rat	bat	cat	hat	spoon	moon	snake	cake	rake
bear	chair	hair	pear	well	bell			

Table 18. Normative names of stimuli used during initial and secondary emergent testing, reinforcement training, and generalisation testing.

Stimuli were 33 black-and-white pictures of easily nameable items (some adapted from Snodgrass & Vanderwart, 1980), the normative names of which were each between three and five letters in length (see Table 18). Sixteen of these stimuli provided the potential for 4 four-member equivalence classes in each condition; another nine, presented in the final testing phase only, the potential for 3 three-member generalised classes. The remaining eight stimuli were only presented during reinforcement training. Equivalence and generalised classes were composed

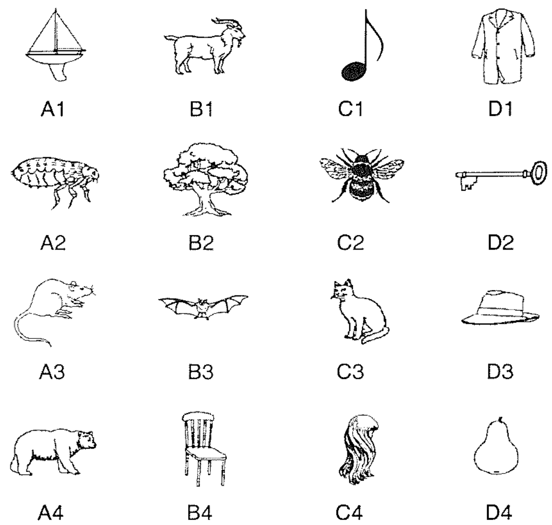

Figure 41. Stimuli and class configurations used in initial and secondary emergent trials for all conditions. Numbered rows denote classes; lettered columns denote stimuli.

of stimuli whose names rhymed with each other, and trials involving these classes always presented a sample whose name rhymed with that of the positive comparison but never with those of the negative comparisons (see Figure 41 and Figure 42 respectively, for stimuli and class arrangements). Reinforcement training

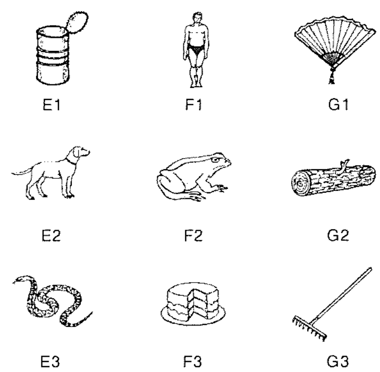

Figure 42. Stimuli and class configurations used in generalisation trials for all conditions. Numbered rows denote classes; lettered columns denote stimuli.

trials always presented a sample whose name rhymed with that of the comparison (see Figure 43).

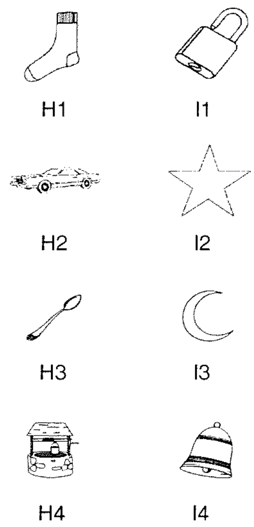

Figure 43. Stimuli and class configurations used during reinforcement training trials for all conditions. Numbered rows denote sample-comparison pairings; lettered columns denote stimuli.

8.2.4 Procedure

Participants were asked to familiarise themselves with the instructions before them, and were then left to complete the experiment.

8.2.4.1 Instructions.

Initially, the following text was displayed on the computer's monitor: "When you are familiar with the written instructions, please click on 'Continue' to start the experiment." The written instructions were as follows:

> When the experiment begins, and at the start of each subsequent trial, you will see a picture in the middle of the screen in front of you. Use the mouse to click on it. One or more other pictures will now appear. Use the mouse to click on one of these. During some stages of the experiment, you may receive feedback on your choices, a "beep" for correct, and a "buzz" for incorrect. For most of the experiment, however, you will not receive feedback on your choices--but keep on going, and continue do the best you can! Please aim to complete the experiment as quickly and accurately as possible. The computer will record your performance throughout, and a message on screen will tell you when the experiment is over. When you are

ready to start, please click on "Continue". Thank you for participating in this experiment. You are free to leave at any point.

The specified action removed the on-screen instructions and match-to-sample training commenced.

8.2.4.2 General procedure and match-to-sample contingencies.

Each trial began with presentation of a sample stimulus, an observing response with the mouse causing the comparison stimulus or stimuli additionally to be displayed. All stimuli remained in view until selection of a comparison caused them immediately to disappear, followed, after a 1-s interval, by presentation of the next trial. Comparison selections made within 0.5 s of presentation had no such consequence, however, and all stimuli remained in view. There was no limit to trial duration and the positions of comparison stimuli varied pseudo-randomly from trial to trial. During reinforcement training, the comparison always shared the same numeric designation as the sample (e.g., H1I1). Throughout emergent and generalisation testing, however, comparisons were always the members of all other stimulus classes sharing the same alphabetic designation (e.g., B1, B2, B3, B4). At no point during the experiment did the location of any comparison stimulus remain constant for more than two consecutive trials, nor did the same sample stimulus appear for more than two trials consecutively. During reinforcement training, comparison selection was always followed by a beep and the word "CORRECT" displayed on the screen. During emergent and generalisation testing, however, the only consequence of a response was presentation of the next trial. The overall procedure for all participants was designed as a series of successive training and testing blocks, the details of which are given below.

8.2.4.3 Phase 1: Reinforcement training.

Initially, reinforcement training was presented, with each of the four relations (i.e., H1I1, H2I2, etc.,) presented in pseudo-random order once every four trials. Participant JM received twelve reinforcement trials, participant KS received four reinforcement trials, and participant CB received a single reinforcement trial (i.e., H2I2). For each participant, presentation of the same number of training trials was repeated subsequent to every third consecutive initial emergent testing block in which criterion was not met. Training was repeated in this way until criterion for initial emergent testing was satisfied, or until the experiment ended consequent to presentation of twelve initial emergent testing blocks (see below) in which criterion had not been met. Table 19 shows all reinforcement training trial configurations used during reinforcement training.

Trained relations	
Sa	Co+
H1	I1
H2	I2
H3	I3
H4	I4

Table 19. Reinforcement training trial configurations, using single-sample and single-comparison displays.

8.2.4.4 Phase 2: Initial emergent testing.

Initial emergent relations								
AB			BC			CD		
Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
A1	B1	B2, B3, B4	B1	C1	C2, C3, C4	C1	D1	D2, D3, D4
A2	B2	B1, B3, B4	B2	C2	C1, C3, C4	C2	D2	D1, D3, D4
A3	B3	B1, B2, B4	B3	C3	C1, C2, C4	C3	D3	D1, D2, D4
A4	B4	B1, B2, B3	B4	C4	C1, C2, C3	C4	D4	D1, D2, D3

Table 20. Initial emergent testing trial configurations, using single-sample and four comparison displays.

Subsequent to initial reinforcement training, all AB, BC, and CD baseline relations were presented without reinforcement in 12-trial blocks, with each relation presented in pseudo-random order once every block. On completion of one trial block with 100% accuracy, secondary emergent testing commenced (Phase 3). If criterion had not been met at the end of 12 initial emergent testing blocks (144 trials), the experiment ended. An on-screen message automatically informed participants of this, and asked them to contact the experimenter. Table 20 shows all emergent trial configurations used during initial emergent testing.

8.2.4.5 Phase 3: Secondary emergent testing.

All possible secondary emergent relations except reflexivity were presented, in pseudo-random order, in a maximum of four 36-trial blocks (see Table 21). Generalised identity-matching repertoires were assumed. If participants satisfied the criterion of 36 of 36 class-consistent responses in any one secondary emergent testing block, generalisation testing commenced (Phase 4). If criterion had not been achieved at the end of two consecutive secondary emergent testing block, however, initial emergent testing was again presented without feedback (Phase 2) until all initial emergent relations were again demonstrated with 100% accuracy over one block of test trials. The final two blocks of secondary emergent testing were then presented. Generalisation testing followed completion of the fourth secondary emergent testing block, regardless of performance.

	Secondary emergent relations											
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
Symmetry												
BA	B1	A1	A2, A3, A4	B2	A2	A1, A3, A4	B3	A3	A1, A2, A4	B4	A4	A1, A2, A3
CB	C1	B1	B2, B3, B4	C2	B2	B1, B3, B4	C3	B3	B1, B2, B4	C4	B4	B1, B2, B3
DC	D1	C1	C2, C3, C4	D2	C2	C1, C3, C4	D3	C3	C1, C2, C4	D4	C4	C1, C2, C3
One-node transitivity												
AC	A1	C1	C2, C3, C4	A2	C2	C1, C3, C4	A3	C3	C1, C2, C4	A4	C4	C1, C2, C3
BD	B1	D1	D2, D3, D4	B2	D2	D1, D3, D4	B3	D3	D1, D2, D4	B4	D4	D1, D2, D3
One-node equivalence												
CA	C1	A1	A2, A3, A4	C2	A2	A1, A3, A4	C3	A3	A1, A2, A4	C4	A4	A1, A2, A3
DB	D1	B1	B2, B3, B4	D2	B2	B1, B3, B4	D3	B3	B1, B2, B4	D4	B4	B1, B2, B3
Two-node transitivity												
AD	A1	D1	D2, D3, D4	A2	D2	D1, D3, D4	A3	D3	D1, D2, D4	A4	D4	D1, D2, D3
Two-node equivalence												
DA	D1	A1	A2, A3, A4	D2	A2	A1, A3, A4	D3	A3	A1, A2, A4	D4	A4	A1, A2, A3

Table 21. Secondary emergent trial configurations, using single-sample and four comparison displays.

8.2.4.6 Phase 4: Generalisation testing.

Two consecutive blocks of 18 test trials each were presented, involving previously unseen stimuli. Each trial presented a novel sample stimulus followed, after an observing response, by three novel comparisons (see Table 22). If, at the end of the first block, participants had selected only comparisons whose names rhymed with those of the samples, match-to-sample testing ended. An on-screen message automatically informed participants of this and asked them to complete the post-test in the envelope before them. Following the second block of generalisation trials, the same message was displayed, regardless of performance.

Generalised emergent relations									
	Sa	Co+	Co-	Sa	Co+	Co-	Sa	Co+	Co-
EF	E1	F1	F2, F3	E2	F2	F1, F3	E3	F3	F1, F2
FE	F1	E1	E2, E3	F2	E2	E1, E3	F3	E3	E1, E2
FG	F1	G1	G2, G3	F2	G2	G1, G3	F3	G3	G1, G2
GF	G1	F1	F2, F3	G2	F2	F1, F3	G3	F3	F1, F2
EG	E1	G1	G2, G3	E2	G2	G1, G3	E3	G3	G1, G2
GE	G1	E1	E2, E3	G2	E2	E1, E3	G3	E3	E1, E2

Table 22. Generalisation trial configurations, using single-sample and three comparison displays.

8.2.4.7 Phase 5: Naming post-test.

Subsequent to match-to-sample testing, participants completed a written post-test that was designed to elucidate their verbal behaviour during the experiment. The booklet was headed by the following instructions:

> Printed below are the pictures that you saw during the experiment. Did you mentally name any of them, or refer to them in any way during testing? If you did, please write beside each picture the name you used for it. If you referred to any picture by more than one name during the experiment, please indicate all the names you used, in the order in which you used them. If you did not refer to a picture in any such way, please leave the space beside it blank. Please note: If the names you used changed during the course of the experiment, you will have opportunity to indicate why they changed at the end of the post-test.

The pictures used during initial and secondary emergent testing were then presented, each followed by a blank space and dotted line. Beneath these were the following instructions:

If you did not see the following pictures, please leave the spaces beside them blank.

Beneath these instructions, the pictures used during reinforcement training were presented, each followed by a blank space and dotted line. Beneath these pictures followed two final questions, each followed by a blank space. For participants JM and KS, the first question was as follows:

> At the beginning of the experiment, you received feedback on the pictures you selected--you were told that your choices were "correct". Did this information lead you to use any mental strategies, "tricks", or rules to decide which pictures went together subsequently, when you no longer received feedback? If it did, please enter a brief description below.

For participant CB, the same question was presented, but the first sentence was as follows:

> At the beginning of the experiment, you received feedback on the first picture you selected--you were told that your choice were "correct".

For all participants, the second question was as follows:

> Did the strategies, "tricks", or rules that you used to decide which pictures went together, or any of the names by which you may have referred to pictures, change during the course of the experiment? If they did, please enter a brief description below, along with any other observations you might wish to make regarding your performance during the experiment.

8.3 RESULTS

All participants completed the experiment. Individual participants' trials and errors during all phases of match-to-sample training and testing are presented in the Appendix I.

8.3.1 Phase 1: Reinforcement Training

All participants completed initial reinforcement training. No participant received additional reinforcement training during the experiment.

8.3.2 Phase 2: Initial Emergent Testing

All participants met the errorless criterion for mastery of initial emergent relations within three testing blocks. Participant CB, who had received a single reinforcement trial prior to initial emergent testing, made three errors during her first testing block and a single error in her second block, but performed errorlessly during her third block of initial emergent testing. Participant KS, who received four reinforcement trials, made four errors during her first initial emergent testing block, but met criterion during her second block. Participant JM, who received twelve reinforcement trials, performed without error during her first block of initial emergent testing.

8.3.3 Phase 3: Secondary Emergent Testing

All participants met the errorless criterion for secondary emergent relations during their first secondary emergent block and proceeded immediately to generalisation testing. No participant received additional initial emergent testing.

8.3.4 Phase 4: Generalisation Testing

All participants selected only rhyme comparisons during the first block of generalisation testing and received no further match-to-sample testing.

8.3.5 Phase 5: Naming Post-test

Although no participant indicated using a non-normative name throughout the experiment, participants JM and CB both indicated that they had not initially referred to *flea* or *tree* by their normative names. All three participants indicated that they had learnt to select comparisons on the basis of rhyme during initial training, and that they had selected comparisons on that basis throughout the rest of the experiment. Participants JM and CB additionally indicated that they had changed to using normative names during the experiment because the non-normative names that they had used initially did not rhyme with the normative names of other stimuli.

8.3.6 Effects of Language Experience

The data from two participants (CY and KY) were rejected because their first language was Chinese. Neither of these participants met criterion for initial emergent testing before the experiment ended after twelve 12-trial testing blocks. Neither of these participants made less than three errors in any initial emergent testing block, and the greatest number of errors made was 11. Despite having received a total of four 12-trial reinforcement training blocks (48 trials), participant CY made an average of 4.4 (SD = 1.4) errors during initial emergent testing, and participant KY made an average of 8.3 errors (SD = 2.7).

8.4 DISCUSSION

The results again suggested that participants named the experimental stimuli without instruction, and that the phonological properties of the names thus given influenced match-to-sample performance. For participant CB, reinforcement training (Phase 1) consisted of a single trial that presented a single comparison stimulus whose normative name rhymed with that of its sample. Participants KS and JM received four and twelve such trials, respectively. For all participants, initial emergent testing (Phase 2) consisted of consecutively presented 12-trial blocks, each composed of novel stimuli, yet by the end of the third such block, all participants had met the 12 of 12 criterion for selection of comparisons whose normative names rhymed with those of their samples. An inverse correlation was indicated between the number of baseline training trials presented to participants and the number of errors made by those participants during initial emergent testing.

Because the stimuli presented during initial emergent testing bore no consistent formal resemblances to each other or to the stimuli presented during reinforcement training, and their names were also phonologically unrelated to those of the latter stimuli, it seems plausible to suggest that participants' selection of comparisons was verbally controlled (Horne & Lowe, 1996), in that during reinforcement training participants had learnt to select comparison stimuli on the basis of rhyme. The uniformly class-congruent selection of rhyme stimuli demonstrated by all participants during the first block of secondary emergent testing (Phase 3) did not undermine this interpretation. Generalisation testing (Phase 4) again presented only novel stimuli that bore no consistent resemblances to any of the stimuli presented previously during the experiment. Nevertheless, during the first testing block, all participants again selected only stimuli whose names rhymed with those of their samples.

Additional confirmation that stimuli were named was provided by participants' written post-tests, which also indicated that the naming of experimental stimuli was almost ubiquitously normative. All participants further indicated that they had learnt to select stimuli on the basis of rhyme during reinforcement training, and had continued to select stimuli on that basis throughout the experiment. Participants JM and CB additionally indicated that the non-normative names by which they had referred to some stimuli had changed during the experiment on the basis of rhyme. Because participant JM selected only rhyme comparisons throughout the experiment, it might further be suggested that on trials involving a non-normatively named stimulus, comparisons had initially been selected by exclusion (i.e., because the normative names of the other comparison stimuli did not rhyme with that of the sample). It is perhaps also of interest to note a correlation between the language experience of two participants exposed to the same training and testing procedure, and their performance during the experiment. The first language of participants CY and KY was Chinese. Despite having received four 12-trial reinforcement training blocks during the experiment, neither participant met criterion for initial emergent testing within twelve testing blocks. It might therefore be suggested that for these participants, the task was not one of simply matching stimuli whose names rhymed.

In summary, therefore, the findings of Experiment One indicate that, among verbally able adults, bidirectional classes of stimuli can emerge in the absence of direct reinforcement and as a result of verbal control previously engendered during reinforced selection of rhyme stimuli. Might such classes be demonstrated to emerge in the absence of *any* reinforcement training, however, and thus in the absence of direct experimentally established verbal control?

8.5 EXPERIMENT TWO: INTRODUCTION

To extend the findings of Experiment One, a second experiment was proposed to investigate whether class-consistent bidirectional relations between stimuli can emerge in the absence of any reinforcement training. Whereas during Experiment One, participants had initially been exposed to reinforcement training involving sample-comparison pairings of easily nameable, yet formally unrelated, stimuli whose names rhymed, no such training was presented prior to testing in extinction during Experiment Two. Although the same stimuli were presented in the same class arrangements as during Experiment One, participants were exposed to reinforcement training only subsequent to repeated failure exclusively to select rhyme comparisons during initial emergent testing. Experiment Two therefore addressed the following question: Can participants' unprompted verbal behaviour alone suffice to establish bidirectional classes of formally unrelated stimuli whose names are phonologically related?

8.6 METHOD

8.6.1 Participants

Twelve students and staff at the University of Southampton (seven female, five male) volunteered to participate in the experiment. Aged between 19 and 36 years, all were native English speakers with no prior knowledge of the research. Participation was voluntary but paid at a rate of £2.50 per 30 min, independent of experimental performance.

8.6.2 Apparatus and Setting

As for Experiment One.

8.6.3 Stimuli and Class Arrangements

As for Experiment One.

8.6.4 Procedure

As during Experiment One, all participants were asked to familiarise themselves with the instructions before them and then left to complete the experiment.

8.6.4.1 Instructions.

As for Experiment One.

8.6.4.2 General procedure and match-to-sample contingencies.

All participants were exposed to the same performance-contingent testing programme as during Experiment One, except that no participant received reinforcement training prior to initial emergent testing.

8.6.4.3 Phase 1: Initial emergent testing.

Initially, AB, BC, and CD baseline relations were presented to all participants without reinforcement in the same pseudo-random order, and to the same criterion, as during Experiment One. For six participants (AM, HR, RC, RM, AF, and IN), initial reinforcement training (Phase 2) was presented if criterion for initial emergent relations had not been met within three testing blocks. For six other participants (AC, PS, MM, PD, RO, and WK), initial training was presented if criterion had not been met within 12 initial emergent testing blocks. Table 23 shows the maximum number of initial emergent testing blocks presented to all participants prior to presentation of initial reinforcement training and the number of trials composing that training, if presented. As during Experiment One, secondary emergent testing (Phase 3) commenced immediately subsequent to completion of one initial emergent testing block with 100% accuracy and the experiment ended if criterion for initial emergent relations had not been met within 12 testing blocks (144 trials) of cessation of initial reinforcement training.

	1 trial	4 trials	12 trials
3 blocks	AM HR	RC RM	AF IN
12 blocks	AC PS	MM PD	RO WK

Table 23. Number of initial emergent testing blocks subsequent to presentation of initial reinforcement training and number of trials composing that training, if presented.

8.6.4.4 Phase 2: Reinforcement training.

Reinforcement training relations were presented in the same pseudo-random order as during Experiment One and, as in that study, the number of trials composing training varied between participants. For four participants (AF, IN, RO, and WK), training was composed of 12 reinforcement trials, and for four other participants (RC, RM, MM, and PD) training was composed of four trials. For four further participants (AM, HR, AC, and PS) training was composed of a single reinforcement trial (i.e., H2I2). As during Experiment One, reinforcement training was repeated subsequent to every third initial emergent testing block in which criterion for initial emergent relations had not been met, until criterion was met or the experiment ended.

8.6.4.5 Phase 3: Secondary emergent testing.

As for Experiment One.

8.6.4.6 Phase 4: Generalisation testing.

As for Experiment One.

8.6.4.7 Phase 5: Naming post-test.

Subsequent to match-to-sample testing, all participants completed a written post-test that was identical to that presented during Experiment One, except that the two final questions were as follows:

> At the beginning of the experiment, you did not receive feedback on the pictures you selected--you were not told which choices were "correct" and which "incorrect". At this stage of the experiment, did you use any mental strategies, "tricks", or rules to decide which pictures went together? If you did, please enter a brief description below.

and

> Later in the experiment, you may have received feedback on certain trials. Did the strategies, "tricks", or rules you used subsequently to select pictures, or any of the names by which you may have referred to pictures, change as a result of this feedback? If they did, please enter a brief description below. If you did not receive feedback on any trials, please leave the space below blank.

8.7 RESULTS

All 12 participants completed the experiment. No participant received more than 12 reinforcement training trials, and six participants completed the experiment without having received any reinforcement training. The trial and error data of participants AM, HR, RC, RM, AF, and IN during all phases of match-to-sample training and testing are presented in the Appendix J, and those of participants AC, PS, MM, PD, RO, and WK appear in Appendix K.

8.7.1 Phases 1 and 2: Initial Emergent Testing and Reinforcement Training

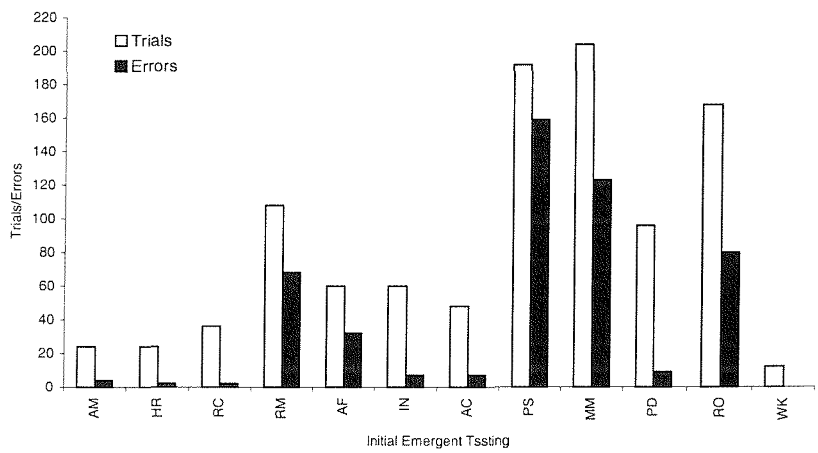

Figure 44. Total trials and errors during initial emergent testing (Phase 1) prior to secondary emergent testing for all participants.

Of the six participants who did not receive reinforcement training prior to meeting the errorless criterion for initial emergent testing, one (WK) performed without error during her first testing block, and two others (AM and HR) performed likewise during their second block. Participant RC met criterion during her third testing block, and two others (AC and PD) performed likewise in their fourth and eighth blocks, respectively. Figure 44 shows the total number of trials required and errors made by all participants during initial emergent testing prior to secondary emergent testing.

Six participants received reinforcement training during the experiment. Of these, AF and IN received twelve reinforcement trials subsequent to their third block of initial emergent testing, and met criterion during their fifth testing block. Participant RM received four reinforcement trials at the same point, another four reinforcement trials subsequent to her sixth initial emergent testing block, and met criterion during her ninth testing block. Participant RO received twelve reinforcement trials subsequent to his twelfth initial emergent testing block and met criterion during his fourteenth block. At the same point, participants MM and PS received four and one reinforcement trials respectively, and both received further training subsequent to their fifteenth initial emergent testing block. MM met criterion during his seventeenth initial emergent testing block and PS performed likewise during his sixteenth block. Figure 45 shows the total number of reinforcement training trials received by all participants prior to secondary emergent testing.

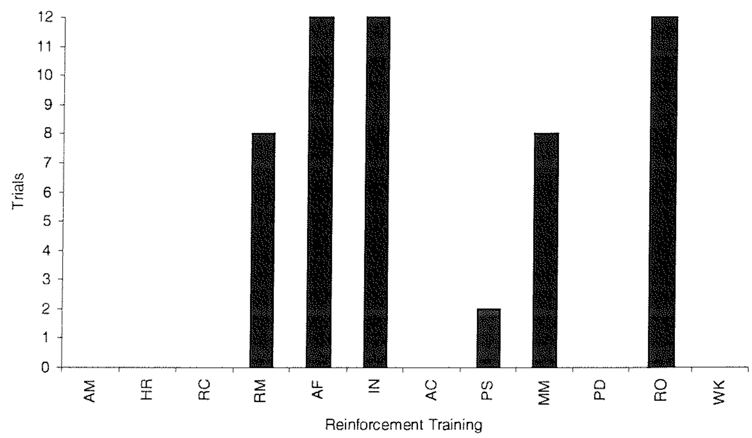

Figure 45. Total trials received during reinforcement training (Phase 2) prior to secondary emergent testing for all participants.

8.7.2 Phase 3: Secondary Emergent Testing

All participants but two met the errorless criterion for secondary emergent relations during the experiment. Figure 46 shows total errors made by all participants during secondary emergent testing.

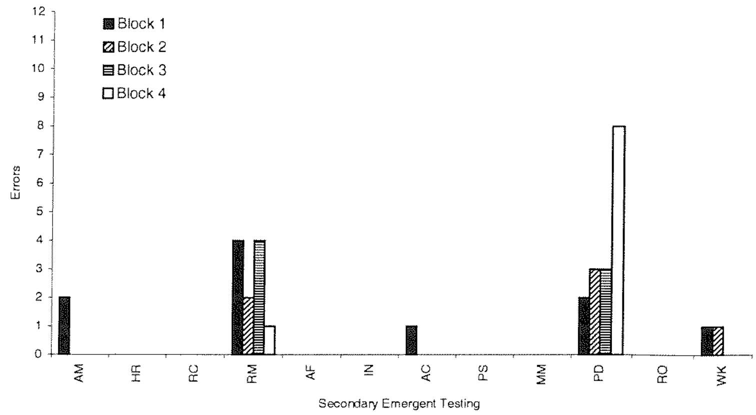

Figure 46. Total errors made by all participants during secondary emergent testing (Phase 3).

8.7.2.1 Secondary emergent testing (blocks 1 and 2).

Of the six participants who did not receive reinforcement training, four (AM, HR, RC, and AC) met criterion within the first two blocks of secondary emergent testing. Of the six participants who received reinforcement training, five (AF, IN, PS, MM, and RO) performed likewise. No participant made more than four errors in either testing block. Participants RM, PD, and WK, who did not meet criterion, made a mean of 3 (SD = 1.4), 2.5 (SD = 0.7), and 1 (SD = 0) errors respectively during secondary emergent testing (blocks 1 and 2).

8.7.2.2 Review prior to secondary emergent testing (blocks 3 and 4).

All three participants who had not met criterion during secondary emergent testing (blocks 1 and 2) performed without error during the first subsequent block of initial emergent testing and proceeded immediately to secondary emergent testing (blocks 3 and 4).

8.7.2.3 Secondary emergent testing (blocks 3 and 4).

Participant WK performed without error during her third block of secondary emergent testing. Participants RM and PD did not meet criterion in either testing block, although RM made only a single error during her fourth block. These participants made, on average, 2.5 (SD = 2.1) and 5.5 (SD = 3.5) errors during secondary emergent testing (blocks 3 and 4), respectively.

8.7.3 Phase 4: Generalisation Testing

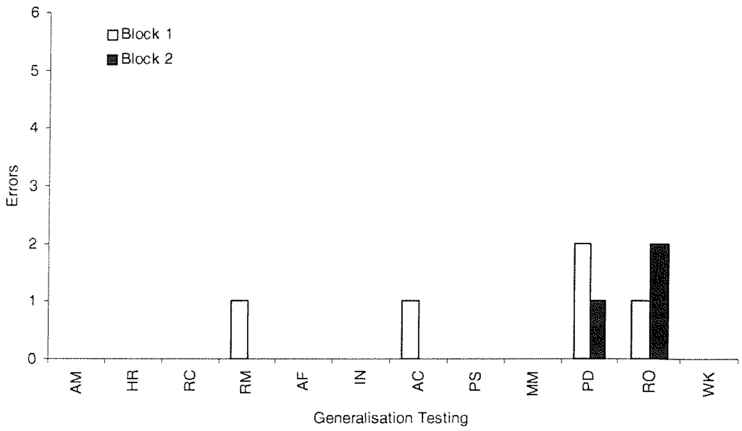

Figure 47. Total errors made by all participants during generalisation testing (Phase 4).

All participants but two met criterion for generalisation testing. Of the six participants who did not receive reinforcement training during the experiment, four (AM, HR, RC, and WK) selected only rhyme comparisons during their first testing block, and AC performed likewise during her second block. PD selected, on average, 1.5 (SD = 0.7) non-rhyme comparisons during generalisation testing. Of the six participants who received reinforcement training, four (AF, IN, PS, and MM) selected only rhyme comparisons during their first testing block, and participant RM performed likewise during her second block. Participant RO selected, on average, 1.5 (SD = 0.7) non-rhyme comparisons during generalisation testing. Figure 47 shows total errors made by all participants during generalisation testing.

8.7.3.1 Phase 5: Naming post-test.

Although no participant indicated using a non-normative stimulus name throughout the experiment, participants' post-tests suggested that some stimuli had not initially been referred to by their normative names. Of the six participants who received reinforcement training during the experiment, two (HR and AC) indicated that they had initially referred to *flea* as either *cockroach*, *louse*, or *bug*, and three participants (HR, AC, and WK) indicated that they had changed from referring to *rat* as *mouse* during the course of the experiment. Likewise, three participants (HR, AC, and WK), indicated initially using *tin* for *can*. Participant HR additionally indicated initially using *fly* for *bee*, and WK noted using *yacht* for *boat*. Of the six participants who received reinforcement training, two (AF, RO) indicated that they had referred to *flea* as either *nit*, *louse*, or *bug*, and three (RM, AF, PS, and RO) indicated that they had changed from using *mouse* to *rat* during the course of the experiment. Two participants (RM and MM) indicated using *tin* for *can,* and participant RO noted using either *music* or *key* for *note*.

All participants indicated that they had selected stimuli on the basis of rhyme during the experiment although PD, the only participant to have received no reinforcement training and not to have met criterion for secondary emergent testing, noted that he had selected stimuli on the basis of other verbal rules (e.g., pictures whose names shared the same first letter) if he "could not locate a rhyming comparison". HR and AC, who also did not receive reinforcement training, indicated that they had changed to using normative stimulus names on the basis of rhyme. Four participants who received reinforcement training (RM, AF, PS, and RO) indicated likewise, and additionally noted the change as a result of reinforcement training. Participant RM further indicated that despite receiving reinforcement training, the images of pairs of stimuli (e.g., *bat in a tree*) that she had formed previously still influenced comparison selection throughout the experiment. Four of the participants who had received training also indicated that they had initially formulated rules as a basis for consistent comparison selection: AF and RO noted the contraction of pairs of stimulus names to a single word (e.g., *key* + *note* = *keynote*) and PS, MM, and RO noted selecting stimuli on the basis of physical similarity (e.g., *bat* and *bee* both fly). Participant MM also indicated initially having used spatio-temporal patterns as a basis for comparison selection.

8.8 DISCUSSION

The results once again suggested that participants had named the experimental stimuli without instruction, and that the phonological properties of the names thus given had influenced match-to-sample performance. For all participants, initial emergent testing (Phase 1) consisted of consecutively presented 12-trial blocks, each composed of novel stimuli. Six participants received reinforcement training if criterion for initial emergent testing had not been met within three testing blocks, and six other participants received training if criterion had not been met within 12 testing blocks.

Six participants met criterion for initial emergent testing in the absence of any reinforcement training. One did so during the first initial emergent testing block, and three others did so within three blocks. Because the stimuli involved bore no consistent formal resemblances to each other, it seems reasonable again to suggest that participants' selection of comparisons was verbally controlled (Horne & Lowe, 1996) or rule-governed (Skinner, 1969), even though selection of rhyme stimuli had not been reinforced during the experiment. The findings of studies One, Two, and Three suggested that relations trained between stimuli can be superseded by verbal control during testing in extinction, if a ready verbal basis for categorisation of stimuli is available. It does not seem implausible to suggest, therefore, that such a verbal basis had once again set the occasion for selection of rhyme comparisons.

Although six participants required reinforcement training to meet criterion for initial emergent testing, no participant received more than 12 reinforcement trials during the experiment. An inverse correlation was again indicated between the number of reinforcement training trials presented to participants and the subsequent performance of those participants during initial emergent testing: All three participants who received a single 12-trial reinforcement training block met criterion for initial emergent testing during their second subsequent testing block. All three participants who received one- or four-trial reinforcement training blocks, however, required two such blocks to meet criterion for initial emergent testing. The behaviour of these participants subsequent to cessation of reinforcement training may again indicate that, for these participants, the verbally controlled selection of rhyme stimuli engendered by training had superseded previous selection of non-rhyme stimuli.

All participants but three met the errorless criterion for secondary emergent relations during their first or second testing blocks, and all but two of those participants did so during their first testing block. One other participant met criterion during the third block of secondary emergent testing. Of the two participants who never met criterion for secondary emergent relations, one had received two 4-trial reinforcement training blocks prior to secondary emergent testing, and the other had received no training. The first of these participants met the errorless criterion for selection of rhyme stimuli during her second generalisation testing block, however, and the other made a total of only three errors across both generalisation testing blocks. One other participant who had received a single 12-trial reinforcement training block prior to secondary emergent testing, but had met criterion during the first secondary emergent block, also made only three errors during generalisation testing. All other participants met criterion during their first or second testing blocks. Because the stimuli presented during generalisation testing bore no consistent formal resemblances to the stimuli presented previously during the experiment, it again seems reasonable to suggest that comparison selection was verbally controlled, in that participants had again simply selected stimuli whose names rhymed.

Additional confirmation that stimuli were named was provided by written post-testing, although this measure indicated greater variability in participants' stimulus naming than reported during Experiment One. Nevertheless, all participants noted selecting stimuli on the basis of rhyme during the experiment. All three participants who received two reinforcement training blocks prior to secondary emergent testing indicated that they had also selected stimuli on the basis of non-verbal strategies (e.g., spatio-temporal location or extra-experimental associations) during the experiment. The single participant who had not received baseline training, but did not meet criterion during secondary emergent testing or generalisation testing, further indicated that he had selected comparisons on the basis of other verbal rules (i.e., stimuli that began with the same initial letter as the sample) during these testing blocks. Because the participant who made no error during initial emergent testing indicated that she had originally referred to *flea* by a non-normative name, it might again be suggested that on trials involving this stimulus, she had initially selected comparisons by exclusion, and that that exclusion had been verbal in nature.

The findings of experiments One and Two therefore indicate that participants name stimuli during experimentation, and that participants' selection of stimuli can be verbally controlled. When the names by which participants refer to stimuli provide a ready verbal basis for stimulus categorisation, participants consistently select stimuli on that basis. When strategies, verbal and otherwise, that are not congruent with that basis are employed, however, consistent bidirectional relations between stimuli do not emerge, or only emerge subsequent to more extensive exposure to the experimental contingencies. These findings necessarily question the description of equivalence classes as arbitrary classes (Stromer & Mackay, 1996). If arbitrary classes are defined as classes composed of finite numbers of formally unrelated stimuli, and equivalence classes as special instances of such classes in that the relations between the stimuli of which they are composed possess the properties of reflexivity, symmetry, and transitivity, then the classes demonstrated during experiments One and Two should be so described. Nevertheless, within the equivalence paradigm, the reflexive, symmetric, and transitive properties of relations between stimuli can only be defined with regard to the reinforcement training by which those stimuli have come to be related. No such training was presented during experiments One and Two, however, and the absence of direct reinforcement rendered the labelling of relations between stimuli by the experimenter essentially arbitrary. Participants' consistent selection of novel stimuli during generalisation testing might further suggest that the emergent classes demonstrated could have been expected potentially to be composed of unlimited numbers of stimuli whose names rhymed with each other.

If the emergent classes demonstrated during experiments One and Two cannot legitimately be described as equivalence classes, by the usual definition they also cannot be described as feature classes: Such classes are defined on the basis of physical similarity between the stimuli of which they are composed (Stromer & Mackay, 1996), yet the emergent classes demonstrated during experiments One and Two were composed of stimuli that were formally unrelated. The results of these experiments may therefore suggest that stimuli can participate in feature classes not only on the basis of physical similarity, but also on the basis of the phonological similarity of the names by which stimuli are referred to. However described, the findings once again indicate that when a verbal basis for classification of stimuli exists, bidirectional classes of stimuli can emerge in the absence of reinforcement training, direct or otherwise.

9. GENERAL DISCUSSION

9.1 INTRODUCTION

This chapter aims firstly to provide a summary of the main findings of the research presented in chapters Four to Eight (Section 9.2), and secondly to address broader theoretical and methodological issues raised by that research (Section 9.3). An evaluation of the three major accounts of untrained human behaviour is provided in Section 9.4, and issues relating to the phonological properties of stimulus names are discussed in Section 9.5. Practical implications of research into emergent behaviour are addressed in Section 9.6, and directions for future research are proposed in Section 9.7. Section 0 concludes the thesis.

9.2 SUMMARY OF MAIN FINDINGS

As Remington (1996) has observed, if participants name stimuli, then the properties of the names given, rather than the stimuli named, should determine the ease with which equivalence classes form. On this basis, it should be possible to identify combinations of stimuli that will be more or less easily related by equivalence on the basis of their names. Study One set out to test this hypothesis by presenting groups of verbally able participants with different arrangements of the same easily nameable, yet formally unrelated, stimuli whose names either did, or did not, rhyme with each other. The results suggested that participants had, without instruction, named the pictorial stimuli to which they were exposed, and that the phonological properties of the names thus given had influenced match-to-sample performance. When the names of stimuli composing classes rhymed with each other, baseline learning and equivalence class formation occurred more quickly and reliably than when those names did not rhyme. Although these findings did not demonstrate the necessity or sufficiency of naming for equivalence and generalised class formation, they nevertheless provided an initial demonstration of the functionality of participants' verbal behaviour during equivalence research.

Study Two set out not only to confirm the findings of Study One using a within-participant design, but also to investigate the potentially functional role of participants' verbal behaviour in the formation of contextually controlled equivalence classes. To achieve this, rhyme and non-rhyme combinations of the same easily nameable, yet formally unrelated, visual stimuli were presented to individual verbally able participants under the control of contextual cues. The results again suggested the functionality of participants' verbal behaviour: Throughout the majority of emergent trials, rhyme stimuli were selected regardless of contextual cue. The principal aim of the experiment had nevertheless been to investigate the functionality of verbal behaviour in the formation of contextually controlled equivalence classes, and contextual control of emergent relations had not been established.

Study Three set out to achieve this control through the implementation of a number of procedural refinements. A mixed experimental design additionally allowed the presentation of further combinations of rhyme and non-rhyme stimuli between groups of participants. The results again suggested the functionality of participants' verbal behaviour during experimentation, and documented the emergence of contextually controlled equivalence classes composed of both rhyme and non-rhyme combinations of stimuli. All participants exposed only to rhyme combinations of stimuli demonstrated equivalence with ease, whereas those participants exposed to both rhyme and non-rhyme combinations of stimuli demonstrated equivalence with less facility: These participants made fewer errors on rhyme trials than on trials of non-rhyme relations. No participant exposed only to non-rhyme combinations of stimuli demonstrated equivalence.

Study Four set out to investigate the functionality of participants' verbal behaviour in the formation of contextually controlled equivalence classes using a think-aloud procedure. All participants were presented with rhyme combinations of easily nameable, yet formally unrelated, pictorial stimuli under the control of one contextual cue, and alliteration combinations of the same stimuli under the control of another contextual cue. Two control conditions were also employed, however: Participants in one of these conditions received instructions to think-aloud, but were only actually required so to do during emergent testing. Participants in the other condition received no such instructions. The results indicated that the use of a think-aloud procedure had been detrimental to participants' performance because of the task demands of the procedure itself, and through the effects of the instructions inherent in its use.

Study Five was proposed to investigate the role of reinforcement training in the formation of equivalence classes composed of easily nameable, yet formally unrelated, stimuli whose names were phonologically related. Experiment One initially exposed verbally able participants to limited reinforcement training involving easily-nameable, yet formally unrelated, pictorial stimuli whose names rhymed with each other. Trials involving different easily nameable, yet formally unrelated, stimuli were presented subsequently in the absence of direct reinforcement. The results once again suggested the functionality of participants' verbal behaviour and further indicated that bidirectional stimulus classes can emerge in the absence of direct reinforcement. A second experiment confirmed these findings, and further suggested that participants can demonstrate the formation of bidirectional stimulus classes in the absence of any prior experimental reinforcement. Although the main findings of studies One to Five thus indicated the functionality of participants' verbal behaviour in equivalence class formation, further theoretical and methodological implications were suggested. These are discussed in Section 9.3.

9.3 FURTHER IMPLICATIONS

9.3.1 Generalisation

An integral element of the research described above was assessment of the potential generalisation of equivalence classes to novel stimuli on the basis of verbal control. Although a number of studies have previously investigated the formation of generalised classes as a function of physical similarity between stimuli (e.g., Adams et al., 1993b; Fields et al., 1996, 1997, 1991; Haring et al., 1989; Meehan & Fields, 1995), no previous research has investigated generalised class formation as a function of participants' verbal behaviour during experimentation.

Generalisation testing during studies One and Five presented participants with novel stimuli that were formally unrelated to the stimuli presented during baseline training and emergent testing, and whose names were also phonologically unrelated to the names of those stimuli. In studies Two and Three, initial generalisation testing was constructed likewise, except that all combinations of novel stimuli were presented in the presence of the contextual cues that had been established to control selection of rhyme and non-rhyme stimuli during baseline training. In all four studies, the name of one comparison stimulus rhymed with that of the sample on every trial. During Study One, all participants who had been trained with rhyme combinations of stimuli consistently selected rhyme stimuli during generalisation testing. It might therefore be argued that for these participants, generalised class formation was the result of verbal control engendered during baseline training. Unexpectedly, however, a number of participants who had been trained with non-rhyme combinations of the same stimuli also consistently selected novel rhyme comparisons during generalisation testing, despite having previously received no reinforcement for such selections. This finding might therefore further indicate that when a ready verbal basis for stimulus categorisation is available during testing without reinforcement, stimuli will be selected on that basis.

Although all participants in Study Two demonstrated mastery of contextually controlled rhyme and non-rhyme baseline relations, the majority of participants selected only rhyme comparisons throughout initial generalisation testing, regardless of contextual cue. Additional indication was therefore provided that contextual control can be superseded by verbal control if a ready verbal basis for stimulus classification is available during testing without reinforcement. All participants who had been exposed only to rhyme combinations of baseline stimuli during Study Three likewise selected only rhyme stimuli during initial generalisation testing, as did one participant who had been trained with contextually controlled rhyme and non-rhyme classes--again indicating the functionality of participants' verbal behaviour during experimentation. Other participants who had received such training during Study Three, however, selected rhyme and non-rhyme stimuli differentially under the control of the contextual cues that had previously controlled selection of rhyme and non-rhyme baseline stimuli. In combination with the findings of Study Two, therefore, these results indicate that the control exerted by contextual cues can generalise on the basis of verbal control: This phenomenon might be conceptualised as "verbal contextual control".

9.3.2 Exclusion

The results of initial generalisation testing during studies Two and Three are also of relevance to the phenomena of exclusion (Dixon, 1977; Green & Saunders, 1998; Johnson & Sidman, 1993; Meehan, 1995; Sidman, 1987; Stikeleather & Sidman, 1990; Stromer, 1986; Stromer & Osborne, 1982; Tomonaga, 1994). Exclusion usually describes selection of stimuli on the basis of reject relations during two choice match-to-sample trials (for example, "if A1 is presented as sample, then the correct comparison will not be B2"). Although, during Study Two, participants had been trained with specific rhyme and non-rhyme sample-comparison pairings, two non-rhyme comparisons and one rhyme comparison were available on every initial generalisation testing trial. That participants selected non-rhyme stimuli under the control of the non-rhyme contextual cue may therefore indicate that exclusion can occur on the basis of verbal control: This phenomenon might be conceptualised as "verbal exclusion".

The notion of verbal exclusion was supported by the results of Study Three. Throughout initial generalisation testing during that study, one rhyme and one non-rhyme stimulus were presented using a two-choice match-to-sample procedure. All participants who had previously been exposed only to rhyme combinations of stimuli selected only rhyme comparisons during initial generalisation testing, as did one participant who had been trained with contextually controlled rhyme and non-rhyme combinations of stimuli. Other participants who had received similar training, however, selected rhyme and non-rhyme stimuli differentially under the control of the contextual cues that had previously been established to control selection of rhyme and non-rhyme stimuli during baseline training. That one participant misnamed *flea* during Study Five, but still selected rhyme only comparisons throughout the experiment appears also to have been on the basis of verbal exclusion. Because there were four comparisons presented on every trial involving this stimulus, exclusion may therefore occur when a greater number of stimuli are employed than has previously been suggested (e.g., Sidman, 1987), when control by exclusion is verbal in nature. In combination with the results of Study Two, therefore, these findings support the utility of concepts of verbal contextual control and verbal exclusion.

9.3.3 Stimulus Compounds

Although investigations into equivalence class formation among elements of complex stimuli may be regarded as a discrete area of equivalence research, potential inter-relationships between control by elements of complex stimuli and the contextual control of equivalence classes have been suggested by some researchers (e.g., Stromer et al., 1993; Sidman, 1994). Previous research has usually employed complex stimuli composed of combinations of two arbitrary auditory or visual stimuli. Secondary generalisation testing during studies Two and

Three, however, aimed to investigate the parameters of verbal contextual control and verbal exclusion using complex stimuli composed of pairs of pictures whose names either rhymed, or did not rhyme, with each other.

The principal aim of secondary generalisation testing during these studies was to assess whether contextual cues that had previously controlled selection of rhyme or non-rhyme stimuli could enter into membership of generalised classes composed of novel multi-element stimuli whose names either did, or did not, rhyme with each other. A number of participants in Study Two selected only rhyme keys during secondary generalisation testing, again indicating that contextual control can be superseded by verbal control during testing without reinforcement. The results of secondary generalisation testing during Study Three supported this conclusion. The results of Study Two additionally indicated, however, that previously established contextual control can generalise to control differential selection of novel stimuli on the basis of participants' verbal behaviour, and that the contextual stimuli exerting that control can themselves enter into membership of generalised classes composed of novel complex stimuli on the basis of the phonological characteristics of the names of the elements composing those stimuli. No published equivalence study has previously reported such findings.

9.3.4 Structure

A number of published studies have, however, reported variations in the speed and accuracy with which participants respond on trials of emergent relations as a function of the nodal distance between stimuli (e.g., Adams et al., 1993a; Bentall et al., 1993, 1999; Dickins et al., 1993; Fields et al., 1993, 1984; Fields & Verhave, 1987; Kennedy, 1991). Studies One to Three set out to investigate potential relationships between participants' verbal behaviour and the effects of nodal distance, using a linear training procedure (cf. Dube et al., 1989; Fields et al., 1990; Kennedy, 1991; Lazar et al., 1984; Lynch & Green, 1991). The results of all three studies indicated that although the errors made by participants on non-rhyme trials of emergent relations were largely consistent with previous reports of the transitivity error effect (Bentall et al., 1993), there were no significant differences in the accuracy with which participants performed on rhyme trials. There was, however, evidence of the transitivity latency effect (Bentall et al., 1993) across both rhyme and non-rhyme trials. Nevertheless, participants' latencies were significantly shorter on rhyme trials than on non-rhyme trials. Horne and Lowe (1996) have suggested that response latencies may reflect participants' precurrent behaviour during experimentation, because such behaviour takes time to emit. Availability of a ready verbal basis for stimulus classification may therefore have led to a reduction in participants' precurrent behaviour, and thus resulted in the shorter latencies observed on rhyme trials. Participants' verbal behaviour should not be disregarded, therefore, in future analyses of the formal structure of equivalence classes.

9.3.5 Instructions and Verbal Reports

Experimental instructions were minimised throughout studies One to Five to avoid contamination of participants' performance by experimenter-induced verbal control (cf. Bush et al., 1989). Nevertheless, "naming may occur... if the stimuli are familiar to subjects" (Stromer & Mackay, 1996, p. 12), and the results of studies One to Five indicated that participants had named stimuli without instruction so to do. As Horne and Lowe (1997, p. 272) have observed, "the verbal repertoire that verbally able human subjects bring with them to an experiment inevitably transforms the experimental environment into one that is also substantially verbal. Whether or not these subjects are provided with experimental instructions, they instruct themselves about their own behavior and its outcomes." The results of studies One to Five endorsed this hypothesis. These studies also employed written post-testing as an indicator of the normativity of participants' naming of stimuli, and the post-testing employed in all studies except Study One additionally assayed the verbal strategies that participants employed during training and testing. As Shimoff (1984, p. 1) has observed, however, "there is no reason to assume that samples of verbal behavior obtained after a session reflect verbal behavior during that session". One might also argue the inverse, however: There is no *a priori* reason to suppose that samples of verbal behavior obtained after a session should *not* reflect verbal behavior during that session--and the results of studies One to Five supported the latter argument. Firstly, all post-tests indicated that participants had named the experimental stimuli, and that they had majoritatively named them normatively. This finding was endorsed by participants' match-to-sample performance and additionally supported by the analysis of participants' concurrent verbal protocols collected during Study Four. As Critchfield et al. (1998) have observed, converging evidence from different types of self-report serves to provide support for the validity of those measures, and the results of Study Four provided such support.

The think-aloud procedure implemented during Study Four produced unexpected results, however. Hayes (1986) and Hayes et al. (1998) have concurred with Ericsson and Simon's (1984, 1993, 1998) assertions that the requirement for participants to verbalise while engaging in a task is not detrimental to task performance, especially when that task is already verbal in nature (Hayes, 1986). The results of Study Four indicated, however, that both thinking aloud and the instructions so to do had been detrimental to participants' performance. This finding was of particular interest because of the difficulties inherent in arguing that the match-to-sample tasks involved had been wholly non-verbal in nature. How might thinking aloud have disrupted performance, however?

Horne and Lowe (1996) have noted that during experimentation, participants' covert verbal behaviour can become contracted in a manner characteristic of "inner speech". Indeed, as Vygotsky (1962, p. 139) has observed, "inner speech appears disconnected and incomplete", and it might therefore be suggested that the requirement for participants to make their covert verbal behaviour overt had disrupted their experimental performance. Nevertheless, on the basis of the verbal behaviour observed during Study Four, one can only concur with Hayes et al.'s (1998, p. 62) statement that "the analysis of verbal protocols may provide a useful contribution to the behavior-analytic study of rule-governed behavior and verbal behavior". More research must be undertaken, however, before final acceptance of Austin and Delaney's (1998, p. 44) assertion that "protocol analysis... can be used to bring critical covert verbal behavior to the overt level", and therefore to clarify the potential relationships between participants' overt and covert verbal behaviour during experimentation.

9.4 EVALUATING THE ACCOUNTS

Sidman and Tailby (1982) proposed the mathematical definition of stimulus equivalence that, together with subsequent theoretical revisions (e.g., Sidman, 1986, 1990, 1992, 1994), has provided the theoretical basis for the overwhelming majority of studies that have investigated emergent behaviour. Although, as Horne and Lowe (1997, p. 271) have pointed out, Sidman has "made a lasting contribution towards ensuring that stimulus classification... and the role it plays in language are put at the top of the research agenda", their "appraisal of the concept of equivalence... remains highly critical" (Horne & Lowe, 1997, p. 271-272).

Verbal and rule-governed behaviour (Skinner, 1957, 1969) are central to Horne and Lowe's (1996) account of emergent behaviour and naming, and these researchers have questioned whether theories of equivalence can account for major aspects of the former phenomena (Horne & Lowe, 1996, 1997). Sidman (1997a, p. 144), has stated that "the equivalence relation can be seen as a rigorous substitute for the popular concept of *correspondence* between words and things [emphasis in original]" and that "the relation between names and the stimuli that occasion them [may be] not only symmetric but reflexive and transitive as well" (Sidman, 1997b, p. 262). Nevertheless, Horne and Lowe (1997, p. 274) have pointed out that the requirement of symmetry between a name and its referent implies logical anomalies: For example, by this requirement, a "child, having learned upon hearing '*where's the boy?*' to look at and point to a boy, should then upon seeing a boy look at and point to the auditory stimulus '*boy*'". The forceful nature of this observation requires little amplification.

Sidman (1997a, pp. 133-134) has nevertheless stated that "mathematical set theory defines the equivalence relation in a way that fits our observations [of behaviour] perfectly. Mathematical abstractions are formulated without reference to real-world specifics, but they are often found to encompass many such specifics." Horne and Lowe (1997, p. 277) have suggested, however, that empirical evidence has already indicated "behavioral relations that should not exist were behavior to follow the orderly 'regularities' of [Sidman's theory]". This suggestion has been supported by research that has reported the repeated non-emergence of equivalence subsequent to protracted training and testing (e.g., Bush et al., 1989; Cowley et al., 1992; Harrison & Green, 1990; Lazar et al., 1984; Lazar & Kotlarchyk, 1986; Saunders et al., 1988; Sidman et al., 1986; Sigurdardottir, Green, & Saunders, 1990; Stromer & Osbourne, 1982). Furthermore, Saunders et al. (1988) have reported that relations of combined symmetry and transitivity can emerge prior to the other prerequisite relations of equivalence, contrary to Sidman's (1994) predictions. Horne and Lowe (1997) have also argued that Sidman's theory fails to account for the substantial differences frequently reported between the behaviour of humans and non-humans, and also for the correlations observed between human participants' verbal abilities and their success on tests of equivalence.

Sidman (1992, 1994) has proposed that equivalence is a behavioural primitive, akin to reinforcement and discrimination in traditional behavioural theory. Owing to data suggesting that both responses and reinforcing stimuli can participate in equivalence classes (Dube et al., 1987, 1989), Sidman (1994, p. 384) has stated that "an equivalence relation is made up of pairs of events, with no restriction on the nature of the events that make up the pairs. The locus of those events, whether it be in the... organism or [its] living or non-living environment, is irrelevant." Because of this, "equivalence relations have their own defining characteristics, none requiring the stimulus/response dichotomy" (Sidman, 1994, p. 386). As Sidman (1994, p. 408) has further observed, this requires that "one large equivalence class *must* emerge when the establishing contingencies share the same reinforcer and defined response [emphasis in original]". As Horne and Lowe (1996) have pointed out, however, not only does this revision remove the distinction between the organism and its environment upon which the majority of psychological theories have been founded, but also highlights the "immense gulf between theory and data" previously noted by Sidman (1994, p. 410).

As Horne and Lowe (1996) have observed, Sidman has placed himself in the invidious position not of having to account for occasional negative findings, but rather of having to explain how the plethora of studies that have employed common reinforcers and defined responses have been able to report the formation of discrete equivalence classes. Although Sidman (1994, p. 408) has attempted to provide explanation in terms of the "selective dropping out" of reinforcers and defined responses from equivalence relations, the principles and conditions controlling such a process, and the process itself, remain to be explained (Horne & Lowe, 1996). As these researchers have further noted, "like many theoretical accounts, Sidman's has evolved to accommodate empirical findings that do not fit easily with his initial formulations". Indeed, "what began as a description of novel behaviors on match-to-sample tasks has... evolved into a very general theory embracing most aspects of behaviour analysis" (Horne & Lowe, 1996, p. 228). These researchers (Horne & Lowe, 1996, p. 230) have pointed out further inconsistencies in Sidman's treatment of verbal rules. Although Sidman (1994) has observed that participants' naming of stimuli may be one means by which the formation of overarching equivalence classes can be prevented--because that naming may provide the conditions of differential responding that allow the formation of discrete equivalence classes--Sidman (1994) has further suggested that verbal behaviour is the result of the prior formation of equivalence classes. The tautology of this explanation is evident.

With regard to Hayes and colleagues' (Hayes, 1991, 1994; Hayes & Hayes, 1989, 1992) account of emergent behaviour, Horne and Lowe (1996) have suggested that relational frame theory is "vague and abstract to the extent that, when not actually misleading, it appears not to be saying very much". Nevertheless, as Barnes (1994, p. 99) has noted, although "relational frame theory may appear somewhat rebarbative at first glance... the basic idea is simple". It aims to provide "an accurate description of the specific type of behavior-environment interactions responsible for equivalence responding and other complex human behaviours" (Clayton & Hayes, 1998, p. 150). With some similarity to Horne and Lowe's (1996) account of emergent behaviour, therefore, "[relational frame theory] explains equivalence as the result of prolonged exposure to contingencies of reinforcement operating within a verbal community" (Clayton & Hayes, 1998, p. 150). Unlike Horne and Lowe's account, however, relational frame theory attempts to explain both equivalence phenomena and verbal behaviour in terms of relational framing (Barnes, 1994).

Horne and Lowe (1996, p. 231) have asked, however, "what is the history" that establishes relational frames and "how does it work?" According to Hayes and Wilson (1996, p. 226), relational frame theory "suggests that relational responses can be brought to bear in the absence of a direct history of relational training with respect to the particular relata". In the words of Hayes and Hayes (1989, p. 174), "it is possible that a child given only a history of symmetry, reflexivity, and transitivity, could derive the frame of co-ordination and show equivalence classes". Although Hayes and Wilson (1996, p. 227) have argued that the quantity and nature of training required for "generalization of a relational response is an empirical matter", Horne and Lowe (1996, p. 231) have argued that appeal to training of this kind "does not clarify the behavioral principles involved in abstracting a frame of co-ordination or sameness from a history of reinforced reversals of stimulus pairs". Although, as Horne and Lowe (1996, p. 233) have pointed out, "relational frame theory acknowledges that naming and verbal rules can help to form relational networks and equivalence" the function of verbal behaviour itself is explained "on the basis of names as contextual cues for relational responses, and on the basis of derived relations formed to the names themselves" (Hayes, 1994, pp. 19-20).

Horne and Lowe's (1996) account, "in many respects a tour de force" (Saunders & Green, 1996), has attracted support among a number of researchers (e.g., Carr & Blackman, 1996, in preparation; Remington, 1996; Stemmer, 1996) but has, in general, been sceptically received (e.g., Chase, 1996; de Rose, 1996; Fields, 1996; Lowenkron, 1996, 1997; Michael, 1996; Moerk, 1997; Peláez-Nogueras, 1996; Whitehurst, 1996). As Chiesa (1992, 1994) has noted, behaviour analysis has traditionally been distinguished by its avoidance of unobservable mediating events or processes, and has looked to behaviour-environment interactions as explanation. As Shimoff (1984, p. 1) has observed, this requires that "the origins of... equivalence relations... do not lie in naming relations", because it would therefore be necessary "that the... complex behavioural repertoire from which equivalence relations [are] inferred [be] presumed to lie in another repertoire... which need not involve overt behaviour" (Stromer & Mackay, 1996, p. 3). Nevertheless, previous research has demonstrated that the development of verbal behaviour greatly alters human operant performance and can account for many of the differences found between human and non-human learning (Bentall et al., 1985)--a proposal originally endorsed by Skinner (1957), who likewise underscored the functional role of unobservable verbal events in the generation of overt human behaviour.

Despite the seemingly conclusive case presented by Horne and Lowe (1996, 1997) for the primacy of verbal behaviour in demonstrations of equivalence and other emergent phenomena, many researchers have remained hostile to their suggestions. A number have also suggested that Horne and Lowe's (1996) account is mediational (e.g., Hayes, 1994; Hayes & Hayes, 1992; Stromer, 1996; Stromer & Mackay, 1996; Stromer, Mackay, & Remington, 1996). For many behaviour analysts, such a term carries pejorative overtones. As Horne and Lowe (1997, p. 273) have pointed out, however, if their account is mediational, then so is much of Skinner's (1957) description of verbal behaviour. For Skinner, "whether one appeals to mediating events, overt or covert, is not an ideological matter but depends on whether the functional analysis demands it". Nevertheless, "a virtue of the naming account is that it dispenses entirely with such mediation" (Horne & Lowe, 1997, p. 274): "The primary role of naming should not be viewed as *mediating* the establish of stimulus classes: Naming *is* stimulus-classifying behavior [emphases in original]" (Horne & Lowe, 1996, pp. 226-227).

Of the three major accounts of emergent relations, Hayes' relational frame theory has perhaps attracted least support from behaviour analysts. Although defended by some (e.g., Barnes, 1994), the failure of relational frame theory to specify the behavioural principles and individual histories governing its operation has been much remarked upon (Horne & Lowe, 1996, 1997; Sidman, 1994). Going little beyond suggestions that a history of reinforced symmetrical responding may be required (Hayes, 1994) with "some small amount of training in combinatorial entailment... probably also... needed" to establish equivalence relations (Hayes & Wilson, 1996), relational frame theory has demonstrated a disconcerting lack of precision and coherence.

Regarding Hayes' (1991) assertion that repeated experience of name-event and event-name reversals is necessary (although not sufficient) for the consequent demonstration of symmetrical arbitrary relations, Sidman (1994) has observed that although stimulus reversals may indeed generate symmetrical relations between formally related stimuli, arbitrary relations (such as reflexivity, symmetry and transitivity) cannot so originate. Although Hayes' premise, if empirically validated, would logically entail the creation of a new behavioural principle (i.e., of learning by repeated stimulus reversals), Sidman (1994, p. 365) has asked "what aspect of several examples of symmetric event-name relations would permit a new example to be recognised or produced" when the exemplars have no distinguishable features in common? In attempting to derive equivalence relations from an individual's behavioural history, therefore, "exemplar theory" does not fulfil its intended purpose; it does not avoid the need to specify a behavioural process that is itself not derivable from anything more basic (Sidman, 1994). Horne and Lowe (1996) have likewise pointed out the contradictory nature of the account, a confusion demonstrated by Hayes' admission that although "'transfer of functions'... instantiates a new behavioural principle", relational frame theory requires the invocation of "no new behavioural principle" (Hayes, 1992, pp. 111-112).

As is evident, debate has been intense--and often acrimonious--in recent years with regard to the respective merits and demerits of the three main accounts of emergent behaviour: But is it possible to reconcile the seemingly disparate? A number of researchers have tried so to do. Barnes (1994, p. 119), for instance, has proposed that "both Sidman's and Hayes's theoretical accounts of stimulus equivalence predict most of the empirical findings in this area of research", and Clayton and Hayes (1998, p. 156-157) have further suggested that "the three main theories share common ground on at least two points. First, all three agree that it is important to establish how subjects, without direct training, can in some contexts treat structurally different stimuli as if they are interchangeable. Second, there is the recognition that the phenomenon of interest is somehow related to linguistic behavior". An emphasis on stimulus classes and classification might be suggested as another area of common ground (cf. Hayes & Barnes, 1997). Clayton and Hayes (1998, p. 158) have noted, however, that they "are left wondering whether any of the theories is more 'true' than any of the others". Perhaps their dilemma might be resolved by consideration of four "features that a good theory must possess" (Sidman, 1997a, p. 125) in relation to the three main accounts of emergent behaviour, and with regard to the findings of the studies that have formed the basis of this thesis? As Barnes (1994, p. 120) has acknowledged, "it is always useful to compare and contrast... empirical accounts".

The first feature that a good theory must possess is "consistency", and Sidman (1997a, p. 141) has proposed that "the consistency of [his] description lies in its replicability. It has been validated in many laboratories and classrooms, with many different kinds of stimuli, with many varieties of subjects, with varying numbers of equivalence classes, with varying class sizes, and with several different teaching and testing procedures". It might also be proposed, however, that Sidman's theory has also proved consistent in its inability to account for repeated failures to demonstrate equivalence, and a substantial amount of other data incompatible with its predictions (Horne & Lowe, 1997). Although it is difficult to assess the consistency of relational frame theory owing to lack of empirical support, Horne and Lowe's (1996) account has provided substantive evidence not only of its consistency with a wealth of published research involving human and non-human participants, but also with Skinner's (1957, 1969) exposition of verbal behaviour and rule-governance. Unlike the former accounts, it also explains with ease the findings of the studies that have formed the basis of this thesis.

The second feature that a good theory must possess is "coherence". As Sidman (1997a, p. 141) has noted, the "criteria for coherence are complex", although "the coherence of [the] descriptive system of [equivalence] lies in its internal cohesiveness" and its "compatibility with other aspects of mathematical set theory". Although the latter assertion is difficult to dispute because the mathematical criteria upon which stimulus equivalence was based were borrowed wholesale from set theory, the former assertion is not so easy to maintain. A number of researchers have already noted the non-equivalence of behavioral and mathematical equivalence (e.g., Holth & Arntzen, 1998; Horne & Lowe, 1996, 1997; Saunders & Green, 1992; Sidman, 1994), not only with regard to theoretical inconsistencies, but also with regard to the behavioural analogue's inability to account for many of the data that it purports so to do. Regarding relational frame theory, it has already been noted that the theory is "vague and abstract to the extent that, when not actually misleading, it appears not to be saying very much" (Horne & Lowe, 1996) and, although supported by some empirical data, it is hard to defend relational frame theory in the terms by which Sidman (1997a) would have his own account appraised. Although Horne and Lowe's (1996) account cannot claim to be coherent with regard to its compatibility with mathematical set theory, this failing has not prevented its authors from proposing an extensive developmental account of naming that coheres with an eclectic selection of multi-disciplinary research and with the findings of the studies that have formed the basis of this thesis.

With regard to "productivity", the third feature of a good theory, it must be conceded that Sidman's account has generated more research than that of either Hayes and colleagues' or Horne and Lowe. Unfortunately, quantity does not necessarily correlate either with utility or with the criteria of consistency and coherence outlined above. Furthermore, a large amount of the empirical evidence generated has already served to undermine the theoretical and ecological validity of Sidman's account and has often, in so doing, further supported the necessity of considering participants' verbal behaviour in any investigation of emergent behaviour (Remington, 1996). That necessity has been further underscored by the findings of the studies that have formed the basis of this thesis.

The final feature by which any theory should be judged is "parsimony". Although Sidman (1997a, p. 142) has stated with regard to his account that "the parsimony is obvious", it might be argued that that parsimony has been eroded in recent years: Not only does Sidman's (1994) account have to explain equivalence in terms of reflexivity, symmetry, and transitivity--it has also to take into consideration concepts of set union, intersection, and partition, and the as yet unexplained process of "selective dropping out". A comparison of the terminology of Hayes and colleagues' account with that of Horne and Lowe may serve to indicate which is the more parsimonious: Relational frame theory explains emergent behaviour in terms of arbitrarily applicable relational responding, which is itself explained in terms of relational frames, which are themselves explained in terms of properties of mutual entailment, combinatorial mutual entailment, and the transfer of stimulus functions. Horne and Lowe would explain such behaviour, including that observed in the studies that have formed the basis of this thesis, with regard to naming.

Although the jury is still out, the weight of evidence suggests a provisional verdict based upon the considerations of consistency, coherence, productivity, and parsimony outlined above: The investigation of verbal behaviour, and naming in particular, will pay dividends greater by far than those to be obtained from extended enquiry into arbitrarily applicable relational responding or stimulus equivalence.

9.5 NAMING AND PHONOLOGY

Studies One to Five set out to demonstrate the functionality of verbal behaviour in the formation and generalisation of equivalence classes. To do this, all participants were trained with combinations of easily nameable, yet formally unrelated, pictorial stimuli whose names either rhymed, or did not rhyme, with each other. The results of all five studies indicated powerful effects of verbal behaviour on the basis of the phonological characteristics of the names by which participants referred to stimuli. When the names of stimuli rhymed with each other, participants demonstrated equivalence and generalised class formation more readily than when those names did not rhyme. But why should this be so?

Goswami and Bryant (1990, p. 3) have noted that "rhyme is an extremely important part of our everyday lives. Rhymes are to be found practically everywhere--in poems, in songs, in advertisements and in political slogans. They are also a significant part of young children's lives. Long before they go to school, they are taught rhymes, and begin to make up their own". A key feature of Horne and Lowe's (1996) account of verbal behaviour has been to chart the development of naming in children during the first two years of their lives. In Horne and Lowe's (1996, p. 227) opinion, "naming is stimulus-classifying behavior", and Goswami and Bryant (1990, p. 22) share a similar view: "To recognise that words rhyme is to put them in categories. These are categories of words which... share a common sound." As these researchers have pointed out, a large body of anecdotal evidence suggests that "children are fascinated by rhyming words from an early age, and use them in their own language games and poems" (Goswami & Bryant, 1990, p. 22). Furthermore, "the experience which children get (often with the active encouragement of their parents) from rhymes... seems to be a natural and spontaneous part of their linguistic development.... When children rhyme they are in effect putting words into categories: these are categories of words with the same end sounds. When children play with alliteration they are also forming categories, this time by the beginning sounds" (Goswami & Bryant, 1990, p. 23). Indeed, "we can be certain that children can detect rhyme and alliteration before they begin to read" (Goswami & Bryant, 1990, p. 23).

It does not seem unreasonable to suggest that this early "phonological awareness" (Goswami & Bryant, 1990) carries through to adult life. Horne and Lowe (1996) have noted that equivalence training and testing often lead to the contraction of the intraverbally linked names employed by subjects (e.g., "*green-cross*" to "*cross*"), and participants' post-tests during studies One to Three supported this observation. These findings might suggest a further behavioural explanation of the facilitative effects of rhyme and alliteration observed in the studies that have formed the basis of this thesis. If, precurrently, participants repeated the names of rhyming and alliterative stimuli during experimentation, the salience of the phonological similarity between those names might have been increased, facilitating stimulus classification and hence class formation. The same behaviour in relation to stimuli with phonologically unrelated names would be unlikely to have such striking effects, as evidenced perhaps by the performance of participants exposed to non-rhyme combinations of stimuli during studies One, Two, and Three. Similarly, from this perspective, exposure to pictures whose names rhyme might have provided a ready basis for the common naming of the stimuli involved (for example, *cat, hat, rat,* and *bat* all share the common phonetic element *-at*). That participants demonstrated equivalence classes of alliterative stimuli with less facility than they demonstrated classes of rhyming stimuli might further be explained by those participants' greater exposure to rhyme during childhood (cf. Goswami & Bryant, 1990) and, indeed, adult life.

9.6 PRACTICAL IMPLICATIONS

All of the "mental" and behavioural phenomena of which psychology is composed are, in a sense, creations of the laboratory. Nevertheless, for an empirical science, experimentation suggests the validity of phenomena and replication lends further endorsement, supporting extrapolation to the "real world". As Sidman (1994, p. 523) has observed, "whenever we find ourselves able to bring a phenomenon under experimental control, turning it on and off at will, the hope is always that we are dealing with something real. Indeed the experience of scientists has taught them that the laboratory *is* part of the real world [emphasis in original]".

The practical implications of stimulus equivalence were suggested from the outset (Sidman, 1971), and subsequent research has appeared both to support and extend that utility. A large number of studies have employed equivalence training and testing to establish reading, writing, and arithmetical skills among learning-disabled participants (e.g., Dube et al., 1992; Joyce & Wolking, 1989; McIlvane et al., 1990; Osborne & Gatch, 1989; Sidman, 1977; Sidman & Cresson, 1973; Stoddard & McIlvane, 1986; Stromer & Mackay, 1992a, 1992b, 1993; Stromer et al., 1996, 1992; Zygmont et al., 1992), and the utility of observational learning techniques in teaching literacy skills among similar participant populations via equivalence has also been reported (Macdonald et al., 1986; Stoddard & McIlvane, 1986). Nevertheless, Sidman (1994, p. 532-533) has noted that "it has been painful for me and my colleagues to experience the widespread disregard for our work on equivalence by school administrators and by the faculties of institutions that are supposed to train the teachers of children.... Applications based on equivalence research have not yet been featured in books for teachers". Perhaps those administrators, publishers, and faculty members have entertained concerns regarding the ecological validity of an approach that places abstract theory at a higher premium than the evaluation of verbal behaviour? Considering Sidman's (1994, p. 532) views regarding "the seemingly mysterious correspondence between the mathematical and behavioral formulations of equivalence relations", they might easily be forgiven.

Compatibility with other areas of psychological enquiry might also be regarded as a practical implication of enquiry into emergent relations and, for a number of researchers (e.g., Barnes, 1996; Barnes & Hampson, 1993; Barnes &

Holmes, 1991), stimulus equivalence represents a potential area of rapprochement with cognitive psychology. The reaction from cognitive psychologists has not been positive, however. Commenting on reactions to his own account, Sidman (1994, p. 265) has noted that "every cognitivist to whom I have shown this structure and described its origins has dismissed it as theoretically uninteresting". Offering commentary to the naming hypothesis, others have been less positive still. Harnad (1996, p. 262), repeatedly assigning the apparently "overlearned" category name "trivial" to Horne and Lowe's (1996) account, has pointed out that it "cannot explain naming capacity". A primer on behaviour analysis might remind him that although such explanations may be the aim of his science, they have always been the antithesis of a functional account. The "associative clusters" and "associative directionality" that Harnad ascribes to the theory are his own mentalism. Nevertheless, considering the majority of behavioural researchers' antipathy towards contemplating the functional role of verbal behaviour in stimulus equivalence, one might feel some sympathy with the views outlined above.

Even if demonstrations of equivalence phenomena in the absence of naming were to become an experimental and applicational certainty, it should lastly be remembered that "the viability of research paradigms is not decided on purely rational grounds" (Hayes & Hayes, 1992). Whitehurst (1996) has suggested that stimulus equivalence may fail the Aunt Sarah test. Your "Aunt Sarah" is intelligent, but unversed in your area of study. Over tea, she asks you what your research is about, and you explain. If your narrative leaves her trying to hide her disbelief as to why anyone should be doing what you're doing, your theory has failed the test. Next time you are round for tea, try to convince your Aunt Sarah that equivalence relations are a behavioural "primitive" which underlies the fact that she knows the sky is blue, and can say as much--and hope that she is not a cognitive psychologist.

9.7 FUTURE DIRECTIONS

A number of directions for future research have been suggested by the studies that have formed the basis of this thesis and by previously published research. Dickins et al. (1993), for instance, demonstrated that participants' verbal behaviour can either facilitate or hinder class formation, depending on the congruence of the naming strategies employed with the experimenter-designated classes governing positive test outcomes. With regard to the facilitative effects of rhyme observed in studies One to Five, it might be possible further to explore the effects of stimulus name phonology (cf. Goswami & Bryant, 1990) in equivalence class formation by pre-training participants with non-normative names for arrangements of easily nameable visual stimuli whose names would otherwise have been phonologically related. If such pre-training could be shown to reduce the facility with which participants demonstrate equivalence, the findings of Dickins et al. (1993) would be extended and an additional demonstration provided of the functionality of participants' verbal behaviour in equivalence class formation. Such research would also allow further investigation of the effects of non-normative stimulus naming indicated by the analysis of participants' verbal protocols during Study Four.

It would also be of interest to explore the inverse of this notion, through the presentation of pre-experimental verbal instructions to participants (cf. Duarte et al., 1998; Eikeseth et al., 1997). A possible basis for such research might be suggested by the indirect rhyme (e.g., *tree-cat*, *bee-hat*, *key-rat*, etc.) available to participants trained with diagonal combinations of stimuli in studies One, Two, and Three. Pre-experimental instructions regarding such phonological similarities might provide participants with a ready verbal basis for stimulus classification, and therefore facilitate the emergence of equivalence between stimuli whose names would otherwise have remained phonologically unrelated. Such a procedure might additionally provide a means by which to differentiate between common and intraverbal naming strategies if a control group was presented with similar arrangements of stimuli without prior instruction.

Another form of experimental instruction could be implemented using a match-to-sample procedure similar to that employed in Study Four. The results of that study indicated that participants demonstrated emergent relations among stimuli whose names were alliterative with less facility than among stimuli whose names rhymed. Contextual control could be provided, and instruction thus given (cf. Lowenkron & Colvin, 1995; Sidman, 1994), by presentation of an auditory or written stimulus of either "*rhyme*" or "*alliteration*" on every trial: The effects of such verbal contextual cues might be expected further to indicate the functionality of common naming in equivalence class formation and to provide further endorsement of the notion of verbal contextual control. It might also be possible further to assess the functionality of participants' naming using procedures similar to those employed in studies One, Two, and Three, but in conjunction with a distractor task (cf. Kennedy & Laitinen, 1988) similar to the "concurrent vocalisation" techniques (e.g., counting out loud, or repeating nonsense syllables) reported by Goswami and Bryant (1990). Implementation of such techniques might be expected to disrupt participants' covert verbal behaviour, and therefore also to disrupt the formation of equivalence relations between stimuli whose names are phonologically related.

Two further issues are raised by the results of studies Four and Five: These studies employed baseline training procedures dissimilar to those used in studies One, Two, and Three and the majority of previous equivalence research (cf. Mandell & Sheen, 1994). Firstly, because significant effects of training order have previously been reported (e.g., Arntzen & Holth, 1997; Saunders et al., 1999, 1988; Smeets et al., 1997; Spradlin & Saunders, 1986; Spradlin et al., 1992; Urcuioli & Zentall, 1993), further investigation of order effects in relation to participants' verbal behaviour would seem warranted. Such investigation could be achieved in a number of ways--for instance, by a replication of Study One using a training technique similar to that employed in studies Four and Five, or by comparing the performance of participants exposed to a training and testing procedure similar to that employed Study One, but additionally using a think-aloud procedure. The results of Study Four also indicate that the effects of think-aloud procedures on participants' verbal behaviour deserve further investigation: This could be achieved by implementing a think-aloud technique similar to that employed in Study Four, but using training and testing procedures similar to those employed in studies One, Two, and Three.

Generalisation testing during studies Two and Three suggested that exclusion can occur on a verbal basis. Because no such finding has previously been reported in the equivalence literature, further investigation of that phenomenon would also seem warranted. This could be achieved, and the parameters of its applicability assessed, by the demonstration of verbal exclusion using an increased number of non-rhyme comparison stimuli. Verbal exclusion on the basis of rhyme might further endorse the validity of that notion if participants could be demonstrated to select specific non-rhyme stimuli in the presence of a number of other comparison stimuli whose names rhymed with that of a given sample. Previous research has indicated that equivalence classes can form as a result of respondent and respondent-type training procedures (e.g., Cullinan et al., 1997;

Leader et al., 1997; Smeets et al., 1997), and it would seem probable that participants would also be able to demonstrate equivalence with facility on the basis of verbal control using such procedures.

The parameters of the functionality of verbal behaviour in the formation and generalisation of equivalence classes could further be investigated with regard to class size. When bidirectional stimulus classes result from verbal control, it has been suggested that the number of stimuli composing such classes may be potentially unlimited. A demonstration of the structural parameters of classes composed of stimuli whose names are phonologically related--especially with regard to the effects of nodal distance--would therefore appear worthwhile. Investigation of the retention and long-term stability (cf. Green et al., 1990; Mackay, 1991; Saunders et al., 1988, 1990) of emergent classes composed of stimuli whose names are phonologically related, in comparison to classes composed of stimuli whose names are phonologically unrelated, would also seem justifiable. Finally, it might be of interest to play Devil's Advocate to the naming hypothesis by investigating the emergence (or non-emergence) of equivalence classes among such easily recognisable, yet notoriously hard to name, stimuli as human faces or chords composed of musical tones.

The directions for future research proposed above address the functionality of verbal behaviour within the overall context of the stimulus equivalence paradigm. Although, for thirty years, equivalence has provided the dominant theoretical framework for the investigation of emergent behaviour, the findings of the studies that have formed the basis of this thesis have questioned the coherence of that theory. Nevertheless, as Horne and Lowe (1997, p. 271) have observed, equivalence has "made a lasting contribution towards ensuring that stimulus classification... and the role it plays in language are put at the top of the research agenda". One can only concur with this statement. One might also suggest, however, that the facilitative effects of rhyme reported in the studies that have formed the basis of this thesis indicate that when the names by which participants refer to stimuli provide a ready basis for stimulus classification, stimuli will consistently be categorised on that basis. One might further propose, therefore, that the role that verbal behaviour plays in the classification of stimuli should also, in future, head the agenda of research into human behaviour, emergent and otherwise.

9.8 CONCLUSIONS

This thesis has presented a programme of research that aimed to investigate the role of verbal behaviour in stimulus equivalence. Primarily, the following question was addressed: Is verbal behaviour functional in the formation and generalisation of equivalence classes and, if so, what are the parameters of that functionality? Overall, the results indicated that participants name stimuli without instruction, and that verbal behaviour is functional both in the formation and generalisation of equivalence classes. Additional research has been suggested further to explore the parameters of that functionality.

Although previous research has addressed the role of verbal behaviour in equivalence, methodological limitations have rendered findings inconclusive. A number of studies have demonstrated correlations between the development of verbal behaviour and success on tests of equivalence (e.g., Devany et al., 1986), but correlation does not necessarily imply causation (Chiesa, 1994). Other studies (e.g., Lowe & Beasty, 1987) have demonstrated that naming interventions can facilitate the emergence of equivalence among participants who had previously tested negative. Mandell and Sheen (1994) have pointed out, however, that such procedures may, despite the use of multiple baseline designs, confound verbal intervention with further exposure to the experimental contingencies. Other research, involving verbally able participants (e.g., Bentall et al., 1993; Dickins et al., 1993; Mandell & Sheen, 1994), may also not have been optimal in that it has relied on differences observed between groups of participants exposed to different stimuli to infer the functionality of naming (Remington, 1996).

Although studies One to Five overcame a number of these limitations, studies One to Four nevertheless indicated neither the necessity nor the sufficiency of verbal behaviour for the demonstration of equivalence. Study Five, however, indicated that verbal control alone may prove sufficient for the formation of bidirectional classes of stimuli although, once again, a demonstration of its necessity was not achieved. Regarding such necessity, and the primacy of equivalence or verbal behaviour, Sidman (1994, p. 567) has observed; "it is a ... 'chicken or egg?' kind of problem. Do equivalence relations help us to explain some aspects of verbal behavior--for instance, the 'specification' of contingencies by rules, or does verbal behavior--for example, rules--make equivalence relations possible? Most people believe the latter... but I am not certain."

On the basis of previous research and the findings of studies One to Five, however, it is not difficult to concur with Horne and Lowe's (1997, p. 272) statement that "until it addresses the role of verbal control, conditional discrimination research with humans will make little significant progress." Indeed, one might also agree with the observation that "if the phenomena of stimulus equivalence do prove to be fully explicable as the product of verbal re-presentation, it is hard to avoid the conclusion that the sustained excitement generated in behavior analysts [regarding stimulus equivalence] has been somewhat misplaced" (Remington, 1996, p. 243). Furthermore, even if the emergent phenomena of equivalence were to prove fully explicable without reference to participants' verbal behaviour, might one not still be justified in asking what is the utility, with regard to the prediction and control of behaviour--or otherwise, of being able to say that stimuli, responses, or any other environmental or behavioural events are simply *equivalent* to one another?

One might further argue that research into stimulus equivalence has aimed, in effect, to achieve a precise definition of one single word in the English language--*equivalence*. With this in mind, one might consider that the New Shorter Oxford English Dictionary (Brown, 1993) contains definitions--albeit somewhat less comprehensive ones--of approximately 300,000 other such English words. Considering that it has required thirty years, and something over three hundred empirical studies, to provide a still intensely disputed definition of one word, to deal likewise with even a tiny proportion of any verbally able human's vocabulary would appear daunting. Might it not therefore be more profitable for research in future to focus on the naming that has filled so many dictionary pages? Verbal behaviour is functional and is the key to human behaviour. Research that ignores that functionality does so at its peril.

APPENDIX A

Study One: Trials (Tr) required and errors (Er) made by all participants (Ps) in the Rhyme (Rh), Orthogonal (Or), and Diagonal (Di) conditions during each stage of match-to-sample training and testing: Phase 1, Baseline establishment (AB, BC, CD); Phase 2, Reinforcement baseline review (Br+); Phase 3, Extinction baseline review (Br-); Phase 4, Emergent testing (Em); Phase 5, Generalisation testing (Gn). Blank cells indicate that no training or testing occurred during that phase. Data rejected from analysis (participant PM) are presented separately.

Ps	AB		BC		CD		Br+		Br-		Em1		Em2		Br-		Br+		Em3		Em4		Gn1		Gn2	
(Rh)	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
JW	32	8	13	1	12	0	12	0	24	1	36	0											18	0		
SG	83	33	25	9	12	0	12	0	12	0	36	7	36	0									18	1	18	0
CS	15	2	12	0	12	0	12	0	12	0	36	0											18	0		
AL	13	1	12	0	12	0	12	0	12	0	36	0											18	0		
HM	31	13	12	0	12	0	12	0	12	0	36	1											18	0		
LL	18	5	12	0	12	0	12	0	12	0	36	0											18	0		
PJ	12	0	12	0	12	0	12	0	24	1	36	0											18	0		
CW	18	4	12	0	12	0	12	0	12	0	36	1	36	1									18	0		
EB	31	14	12	0	12	0	12	0	12	0	36	0											18	1	18	0
MH	20	8	12	0	21	1	12	0	12	0	36	0											18	0		

Ps	AB		BC		CD		Br+		Br-		Em1		Em2		Br-		Br+		Em3		Em4		Gn1		Gn2	
(Or)	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
JH	26	7	22	8	55	18	48	6	36	2	36	20	36	21	24	1			36	19	36	16	18	12	18	6
MG	61	31	44	22	60	32	36	2	12	0	36	21	36	18	12	0			36	18	36	17	18	8	18	14
MB	14	1	44	18	29	10	72	10	12	0	36	18	36	4	36	2	24	1	36	4	36	5	18	11	18	7
CN	29	12	54	18	36	15	108	24	12	0	36	4	36	5	12	0			36	1			18	0		
SS	57	34	24	5	27	9	96	11	12	0	36	22	36	13	36	2	24	1	36	6	36	5	18	2	18	0
JS	25	7	32	11	62	20	96	13	12	0	36	7	36	3	60	7	72	6	36	4	36	4	18	3	18	0
IP	51	21	44	18	58	23	72	8	24	1	36	3	36	0									18	1	18	0
GC	24	4	29	11	43	14	12	0	12	0	36	16	36	18	24	1			36	17	36	16	18	7	18	0
RH	43	13	20	7	33	5	12	0	12	0	36	11	36	6	24	1			36	0			18	4	18	0
NR	40	10	26	3	45	19	12	0	12	0	36	0											18	8	18	0

Ps	AB		BC		CD		Br+		Br-		Em1		Em2		Br-		Br+		Em3		Em4		Gn1		Gn2	
(Di)	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
JS	27	9	22	8	27	10	168	19	12	0	36	17	36	21	12	0			36	22	36	22	18	11	18	13
KS	48	19	96	37	14	2	108	16	12	0	36	17	36	16	12	0			36	22	36	19	18	14	18	11
JT	61	27	79	30	21	4	348	40	84	13	36	9	36	21	36	3	12	0	36	20	36	23	18	15	18	15
PH	53	17	30	7	40	17	96	13	36	7	36	21	36	22	24	1			36	18	36	22	18	16	18	17
RC	33	13	19	7	15	1	12	0	12	0	36	29	36	36	48	28	48	4	36	36	36	36	18	0		
CS	24	4	45	8	12	0	48	5	12	0	36	23	36	24	36	2	12	0	36	13	36	19	18	2	18	0
MY	35	10	25	5	20	5	48	6	12	0	36	31	36	32	12	0			36	32	36	32	18	18	18	18
CB	16	2	33	11	21	9	60	8	12	0	36	24	36	23	36	5	48	3	36	21	36	25	18	18	18	17
WS	20	2	51	23	18	3	60	6	12	0	36	5	36	6	12	0			36	3	36	3	18	1	18	0
KP	19	7	32	9	24	6	24	1	12	0	36	1											18	2	18	0

PM	37	15	27	13	32	10	12	0	12	0	36	16	36	20	24	1			36	21	36	18	18	13	18	14

APPENDIX B

Study One: Mean response latencies (in seconds) of all participants (Ps) in the Rhyme (Rh), Orthogonal (Or), and Diagonal (Di) conditions on trials of symmetry, one-node transitivity (1-node tr), one-node equivalence (1-nd eq), two-node transitivity (2-nd tr), and two-node equivalence (2-nd eq) during the first block of emergent testing. Data rejected from analysis (participant PM) are presented separately.

Ps	Mean Response Latencies				
(Rh)	Symmetry	1-node tr	1-node eq	2-node tr	2-node eq
JW	2.08	2.76	2.31	2.86	3.43
CS	1.22	1.64	2.26	1.52	1.87
AL	1.53	1.57	1.76	1.7	1.81
HM	2.45	3.34	3.54	2.57	3.85
LL	1.55	1.4	1.4	1.13	2.24
PJ	2.83	2.45	3.07	2.53	4.14
EB	1.64	1.98	2.22	2.64	2.15
MH	1.6	1.97	1.62	1.31	4.01
SG	2.66	2.86	3.53	3.78	5.66
CW	1.98	2.05	2.66	2.06	2.2

(Or)	Symmetry	1-node tr	1-node eq	2-node tr	2-node eq
NR	2.69	5.02	3.87	4.12	6.41
IP	4.46	4.44	4.85	9.78	9.29
CN	4.64	5.91	6.07	7.91	11.52
RH	3.12	5.36	5.69	9.17	6.86
JH	5.21	5.72	5.82	5.18	5.67
MG	2.48	2.8	3.91	2.86	2.29
MB	3.19	6.01	7.88	5.07	8.55
SS	12.93	7.56	18.75	24.07	39.76
JS	2.14	1.77	2.33	2.75	2.87
GC	2.33	5.28	6.25	7.64	6.46

(Di)	Symmetry	1-node tr	1-node eq	2-node tr	2-node eq
KP	5.36	4.67	5.8	7.35	9.26
JS	3.66	8.7	8.06	8.32	10.31
KS	3.46	6.8	6.61	3.21	8.49
JT	2.98	5.52	2.8	2.48	6.83
PH	4.93	6.09	5.86	5.74	6.36
RC	3.51	4.77	4.6	5.77	4.2
CS	3.21	5.82	3.81	4.54	6.26
MY	8.38	7.12	5.29	5.06	11.86
CB	2.33	3.4	2.46	2.79	3.77
WS	3.3	4.99	5.29	9.12	7.94

PM	2.73	4.48	4.7	5.14	7.36

APPENDIX C

Study Two: Trials (Tr) required and errors (Er) made by all participants (Ps) during each stage of match-to-sample training and testing (adjusted for counterbalancing): Phase 1, Rhyme baseline training (R); Phase 2, Reinforcement rhyme baseline reviews (Rbr+); Phase 3, Extinction rhyme baseline reviews (Rbr-); Phase 4, Non-rhyme baseline training (N); Phase 5, Reinforcement non-rhyme baseline reviews (Nbr+); Phase 6, Extinction non-rhyme baseline reviews (Nbr-); Phase 7, Reinforcement and Extinction full baseline reviews (Fbr+, Fbr-); Phase 8, Emergent testing (Emgt); Phase 9, Initial generalisation testing (1Gn); Phase 10, Secondary generalisation testing (2Gn). Blank cells indicate that no training or testing occurred during that phase.

	R1		R2		R3		Rbr+		Rbr-		N1		N2		N3		Nbr+		Nbr-	
Pp	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
AA1	15	2	12	0	12	0	12	0	12	0	41	13	45	13	24	7	36	3	12	0
JW1	25	8	17	5	12	0	12	0	12	0	37	13	36	16	32	7	84	14	24	1
FT	12	0	12	0	12	0	12	0	12	0	33	10	22	6	13	1	48	3	24	1
HL	12	0	12	0	12	0	12	0	12	0	41	10	32	13	33	12	108	13	12	0
MS	20	4	12	0	12	0	12	0	12	0	21	7	24	6	20	4	96	14	12	0
JW2	12	0	12	0	12	0	12	0	12	0	58	11	47	15	12	0	12	0	12	0
AS	20	7	12	0	12	0	12	0	12	0	114	62	35	12	21	5	72	6	12	0
AH	17	1	14	1	12	0	12	0	12	0	20	4	70	25	31	9	72	8	12	0
RR	31	9	23	5	13	1	12	0	12	0	33	12	25	10	13	1	48	7	12	0
JP	19	7	12	0	12	0	12	0	12	0	61	25	43	7	23	7	120	17	12	0
RW	27	7	13	1	12	0	24	1	12	0	37	14	26	9	17	3	24	2	12	0
AW	12	1	12	0	12	0	12	0	12	0	68	19	30	8	16	3	24	0	24	1
AA2	24	9	12	0	12	0	12	0	12	0	20	4	24	7	24	9	48	5	12	0
AJ	29	17	12	0	12	0	12	0	12	0	28	10	57	21	20	6	48	4	12	0
AC	22	4	12	0	12	0	12	0	12	0	31	8	31	10	20	6	36	2	12	0
CB	19	4	12	0	12	0	12	0	12	0	12	0	21	6	21	7	24	1	12	0

	Fbr-		Fbr+		Emgt 1		Fbr-		Fbr+		Emgt 2		1Gn		2Gn1		2Gn2	
Pp	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
AA1	48	4	0	24	72	35	24	1			72	35	36	2	12	10	12	10
JW1	48	13	24	4	72	32	48	3	24	0	72	36	36	18	12	0		
FT	48	12	48	8	72	34	48	9	24	1	72	38	36	0	12	2	12	0
HL	48	14	48	10	72	34	48	9	24	2	72	36	36	0	12	8	12	6
MS	48	12	48	8	72	34	24	1			72	37	36	0	12	4	12	6
JW2	48	13	48	13	72	27	48	8	24	1	72	25	36	10	12	4	12	1
AS	48	12	48	13	72	34	48	12	24	1	72	37	36	0	12	4	12	5
AH	48	12	48	8	72	34	48	5	24	2	72	30	36	1	12	2	12	0
RR	48	12	48	5	72	17	24	2			72	21	36	0	12	0	12	0
JP	24	2			72	35	48	12	24	2	72	32	36	0	12	8	12	4
RW	24	0			72	25	24	2			72	26	36	17	12	7	12	7
AW	24	2			72	27	24	2			72	26	36	17	12	0		
AA2	24	0			72	30	24	2			72	26	36	1	12	0		
AJ	48	3	48	5	72	26	48	4	24	2	72	29	36	0	12	7	12	5
AC	48	14	72	25	72	36	48	4	48	4	72	33	36	0	12	4	12	4
CB	48	13	48	12	72	32	48	4	48	8	72	33	36	0	12	5	12	5

APPENDIX D

Study Two: Mean response latencies (in seconds) of all participants on trials of symmetry (Sym), one-node transitivity (1-nd tr), one-node equivalence (1-nd eq), two-node transitivity (2-nd tr), and two-node equivalence (2-nd eq) during emergent testing (blocks 1 and 2).

Block 1	Rhyme					Non-rhyme				
Ps	Sym	1-nd tr	1-nd eq	2-nd tr	2-nd eq	Sym	1-nd tr	1-nd eq	2-nd tr	2-nd eq
AA1	2.82	2.20	2.17	1.77	4.47	4.39	5.56	4.84	4.29	4.25
JW1	1.76	2.65	2.87	2.02	3.82	2.06	2.87	3.11	2.52	2.27
FT	2.90	2.42	3.95	2.45	5.30	4.22	3.63	4.10	4.84	5.57
HL	2.41	3.95	2.69	3.03	5.10	2.22	3.80	4.41	1.63	6.76
MS	3.27	4.20	4.31	2.99	3.21	3.01	3.60	4.07	4.13	2.51
JW2	2.04	2.99	2.29	2.83	3.72	3.38	3.11	4.45	6.25	4.75
AT	2.57	4.97	3.39	2.88	3.76	3.56	4.00	3.14	8.22	4.56
AH	3.68	4.07	5.31	3.27	6.15	4.69	6.10	6.21	5.52	9.90
RR	2.18	2.11	2.40	2.15	3.80	2.92	4.40	5.18	5.59	6.08
JP	2.72	2.41	3.10	2.64	2.53	3.16	3.18	2.32	1.70	6.53
RW	2.11	2.17	2.18	1.61	2.58	3.00	4.70	4.81	4.92	4.09
AW	1.65	1.58	2.10	2.33	2.26	4.54	9.85	9.18	8.95	8.02
AA2	2.80	2.06	2.55	2.24	2.38	5.45	7.75	9.60	6.04	9.98
AJ	2.66	2.62	2.51	2.76	3.11	3.65	4.96	6.97	6.47	5.55
AC	5.49	4.30	3.72	8.17	6.31	6.50	7.79	7.08	6.76	9.34
CB	3.83	3.98	2.65	2.02	3.46	5.88	6.87	4.92	3.29	4.91

Block 2	Rhyme					Non-rhyme				
Ps	Sym	1-nd tr	1-nd eq	2-nd tr	2-nd eq	Sym	1-nd tr	1-nd eq	2-nd tr	2-nd eq
AA1	2.09	1.75	2.21	1.48	1.74	3.70	4.59	3.71	3.25	4.07
JW1	1.65	1.62	2.29	1.74	2.54	2.39	2.31	2.14	1.86	2.97
FT	3.55	1.69	2.93	4.23	1.66	2.69	3.59	3.52	2.84	4.30
HL	3.48	2.58	3.05	3.38	2.57	3.08	3.38	3.20	2.32	3.34
MS	3.06	3.52	2.67	2.81	1.80	2.58	3.05	3.05	1.83	3.04
JW2	3.05	2.12	2.96	3.19	2.27	2.91	2.12	2.33	2.09	4.38
AT	3.55	3.57	3.62	4.68	3.65	3.52	4.38	4.06	2.44	4.49
AH	4.24	3.90	6.23	4.30	5.71	4.48	5.66	6.10	5.20	6.25
RR	1.87	1.97	1.85	2.18	3.23	2.13	3.39	3.14	4.70	4.28
JP	2.34	3.24	2.13	2.93	3.48	5.19	2.59	4.40	10.62	4.88
RW	2.08	2.72	2.22	2.11	2.02	3.17	5.15	5.13	4.77	4.31
AW	1.70	1.79	2.28	1.70	3.23	3.14	5.18	6.82	7.88	8.19
AA2	2.41	2.84	2.29	2.03	2.17	6.15	10.03	8.47	9.22	6.78
AJ	3.10	2.60	1.79	3.72	2.71	3.03	2.81	4.59	2.77	1.72
AC	3.06	3.40	3.32	7.86	4.88	6.89	8.13	7.72	6.60	6.62
CB	3.22	3.78	3.33	2.38	4.55	5.37	4.67	5.72	5.88	5.53

APPENDIX E

Study Three: Trials (Tr) required and errors (Er) made by all participants (Ps) in the R/R, R/O, R/D, and D/O conditions (Cs) during each stage of match-to-sample training and testing: Phases 1 and 4, Baseline establishment (AB, BC and EF, FG); Phases 2 and 5, Reinforcement baseline reviews (Cbr+); Phases 3 and 6, Extinction baseline reviews (Cbr-); Phase 7, Reinforcement and Extinction full baseline reviews (Fbr+, Fbr-); Phase 8, Emergent testing (Emgt); Phase 9, Initial generalisation testing (1Gn); Phase 10, Secondary generalisation testing (2Gn). Blank cells indicate that no training or testing occurred during that phase.

		AB		BC		Cbr+		Cbr-		EF		FG		Cbr+		Cbr-		Fbr-		Fbr+		Emgt1		Emgt2		Fbr-		Fbr+		Emgt3		Emgt4	
Cs	Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
R/R	DS1	15	2	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0			24	0										
	JD	14	2	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0			24	0										
	JP	16	4	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0			24	0										
	KM	26	10	15	2	12	0	12	0	12	0	12	0	12	0	12	0	12	0			24	2	24	0								
	MR	13	1	12	0	12	0	12	0	12	0	13	1	12	0	12	0	12	0			24	0										
	TJ	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0	12	0			24	0										
R/O	DS2	15	3	12	0	12	0	12	0	30	13	24	6	48	10	12	0	12	0			24	6	24	5	24	1	24	1	24	7	24	7
	JS	14	2	12	0	12	0	12	0	36	12	18	4	48	6	12	0	12	0			24	1										
	MH	15	2	12	0	12	0	12	0	19	2	25	7	36	4	12	0	72	7	168	12	24	8	24	9	36	2	36	1	24	9	24	8
	OR	14	2	12	0	12	0	12	0	19	2	25	8	24	1	12	0	12	0			24	9	24	8	36	2	72	8	24	7	24	2
	SJ	15	3	12	0	12	0	12	0	32	8	14	1	12	0	12	0	24	1	24	1	24	6	24	6	12	0			24	4	24	4
	SH1	22	2	12	0	12	0	12	0	22	6	28	10	12	0	12	0	12	0			24	7	24	7	24	2	132	25	24	3	24	2
R/D	AF	15	3	12	0	12	0	12	0	37	11	48	17	12	0	12	0	24	6	144	36	24	0										
	BL	14	2	12	0	12	0	12	0	33	11	31	6	24	1	12	0	24	3	24	3	24	6	24	8	24	1	36	2	24	9	24	8
	CP	16	3	12	0	12	0	12	0	18	4	13	1	12	0	12	0	24	6	24	4	24	7	24	8	24	3	24	1	24	4	24	8
	KP	14	2	12	0	12	0	12	0	28	9	47	8	96	9	12	0	12	0			24	5	24	4	24	1	12	0	24	3	24	2
	MP	30	8	14	2	12	0	12	0	15	3	22	5	72	1	12	0	12	0			24	11	24	10	36	3	48	3	24	5	24	2
	RS	13	1	12	0	12	0	12	0	21	6	53	17	36	6	12	0	12	0			24	7	24	8	24	1	24	1	24	7	24	7
D/O	LA	25	6	29	5	24	1	12	0	20	7	25	4	84	10	24	1	24	3	132	28	24	13	24	12	36	2	156	22	24	12	24	13
	NM	17	4	36	9	24	0	24	1	17	3	16	3	24	1	12	0	12	0			24	10	24	11	36	2	96	14	24	11	24	3
	SF	32	8	17	2	48	4	12	0	19	4	28	7	72	13	12	0	12	0			24	6	24	14	12	2	144	27	24	10	24	11
	SH2	45	20	15	2	36	2	12	0	36	6	26	4	24	1	12	0	36	2	48	4	24	12	24	14	24	1	48	5	24	12	24	13
	SM	15	3	18	2	36	7	12	0	17	2	13	1	24	1	12	0	24	1	108	11	24	9	24	12	24	1	12	0	24	14	24	12
	VA	20	6	14	2	72	6	12	0	14	2	17	3	24	1	12	0	12	0			24	10	24	13	36	7	216	39	24	11	24	15

		Fbr-		Fbr+		Emgt5		Emgt6		Fbr-		Fbr+		Emgt7		Emgt8		1Gn1		1Gn2		2Gn1a		2Gn2a		2Gn3a		2Gn1b		2Gn2b		2Gn3b	
Cs	Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
R/R	DS1																	24	0			12	5	12	4	12	2	12	1	12	1	12	0
	JD																	24	0			12	3	12	1	12	0	12	1	12	3		
	JP																	24	0			12	11	12	11	12	1	12	10	12	11	12	1
	KM																	24	0			12	10	12	12	12	0	12	11	12	12		
	MR																	24	0			12	6	12	11	12	1	12	12	12	11	12	1
	TJ																	24	0			12	7	12	12	12	0	12	12	12	12		
R/O	DS2	12	0			24	6	24	9	12	0			24	3	24	5	24	12	24	12	12	1	12	0	12	0	12	0				
	JS																	24	2	24	0	12	1	12	0	12	0	12	0				
	MH	12	0			24	9	24	8	24	2	12	0	24	8	24	6	24	13	24	12	12	9	12	4	12	3	12	6	12	6	12	1
	OR	12	0			24	1											24	4	24	3	12	1	12	1	12	2	12	1	12	2	12	0
	SJ	12	0			24	4	24	5	12	0			24	4	24	5	24	1			12	8	12	6	12	9	12	5	12	1	12	2
	SH1	24	1	24	1	24	3	24	1									24	2	24	2	12	0	12	3	12	1			12	1	12	0
R/D	AF																	24	0			12	8	12	7	12	5	12	8	12	10	12	8
	BL	24	1	12	0	24	8	24	6	12	0			24	7	24	6	24	6	24	6	12	0	12	0	12	0						
	CP	12	0			24	2	24	7	24	1	12	0	24	1			24	2	24	2	12	0	12	0	12	0						
	KP	24	3	36	2	24	6	24	1									24	1			12	4	12	0	12	1	12	1			12	0
	MP	24	2	12	0	24	0											24	0			12	1	12	0	12	0	12	0				
	RS	12	0			24	8	24	9	12	0			24	8	24	7	24	0			12	0	12	0	12	0						
D/O	LA	12	0			24	12	24	16	24	1	192	22	24	12	24	14	24	13	24	13	12	7	12	6	12	7	12	7	12	5	12	6
	NM	24	1	24	2	24	12	24	10	24	1	12	0	24	10	24	8	24	22	24	24	12	11	12	12	12	0	12	12	12	12		
	SF	24	2	132	120	24	8	24	11	36	4	144	14	24	5	24	5	24	12	24	2	12	5	12	5	12	8	12	4	12	6	12	9
	SH2	24	1	132	14	24	12	24	14	24	2	12	0	24	15	24	11	24	16	24	16	12	6	12	6	12	6	12	6	12	7	12	7
	SM	12	0			24	13	24	12	24	1	12	0	24	16	24	14	24	24	24	24	12	11	12	12	12	0	12	12	12	12		
	VA	24	2	48	4	24	11	24	14	24	1	36	2	24	13	24	9	24	24	24	24	12	7	12	6	12	4	12	8	12	4	12	7

APPENDIX F

Study Three: Mean response latencies (in seconds) of all participants (Ps) in the R/R, R/O, R/D, and D/O conditions (Cs) on red and blue trials of symmetry (Sym), transitivity (Tr), and equivalence (Eq) during emergent testing. Blank cells indicate that no training or testing occurred during that phase.

		Block 1						Block 2						Block 3						Block 4					
		Red			Blue			Red			Blue			Red			Blue			Red			Blue		
Cs	Ps	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq
R/R	DS1	1.45	1.46	1.55	1.28	1.44	1.74																		
	JD	1.44	1.11	1.36	1.11	1.01	1.29																		
	JP	1.68	1.62	1.48	1.13	1.21	1.29																		
	KM	1.41	1.54	2.94	1.62	1.83	1.60	1.38	1.72	1.81	1.49	2.59	1.46												
	MR	1.15	1.27	1.38	1.06	1.08	1.32																		
	TJ	1.07	1.19	1.45	0.96	1.02	1.32																		
R/O	DS2	2.77	2.69	3.84	7.69	13.56	6.39	2.65	2.97	2.51	8.78	12.85	12.47	2.99	2.47	4.93	5.09	8.16	5.10	3.14	1.40	5.15	4.29	15.14	8.16
	JS	1.78	1.95	1.83	4.63	6.87	5.86																		
	MH	2.50	2.36	4.03	4.48	5.10	5.19	2.11	3.16	1.27	3.25	3.84	4.21	1.72	2.88	1.83	2.14	5.35	5.33	1.78	3.75	1.98	3.34	4.53	2.47
	OR	1.61	1.68	2.73	2.99	4.01	5.39	1.22	1.89	1.01	2.58	2.56	3.80	1.08	2.76	1.94	2.26	5.53	4.45	0.87	1.56	4.31	2.29	2.00	3.13
	SJ	4.37	1.99	1.97	5.08	6.53	7.21	3.08	2.37	1.60	3.34	5.41	3.05	2.01	2.77	2.67	2.07	4.17	2.54	1.88	3.10	3.68	2.07	3.10	2.77
	SH1	2.86	4.03	5.48	4.53	6.30	8.21	3.24	2.83	3.33	3.17	5.15	7.94	2.83	4.00	1.63	4.46	5.14	7.01	2.68	2.55	2.71	4.09	3.33	3.93
R/D	AF	1.77	2.09	1.73	2.42	2.84	3.56																		
	BL	4.25	1.83	4.13	6.72	9.76	9.60	3.92	3.34	4.42	3.48	7.23	9.96	2.45	1.78	3.23	7.67	9.30	4.95	2.91	3.22	2.37	2.25	8.01	5.55
	CP	2.24	1.46	2.13	3.12	9.30	4.22	1.87	1.39	1.07	1.67	4.15	3.12	1.64	2.32	1.33	2.16	9.62	11.85	1.24	1.57	1.80	2.24	2.41	3.83
	KP	2.46	1.77	1.58	5.27	8.84	8.80	1.47	1.60	1.63	6.52	9.61	3.04	1.74	1.59	1.69	2.05	5.21	5.54	1.48	3.03	3.27	3.32	5.59	5.94
	MP	4.54	5.36	1.42	5.38	2.21	2.06	2.13	1.34	2.61	1.51	1.69	1.29	7.82	7.21	9.04	8.61	5.46	8.29	7.14	6.31	5.36	6.22	8.92	4.91
	RS	4.02	1.60	1.21	4.34	9.67	3.85	3.64	2.26	2.48	3.54	2.89	13.92	1.46	1.09	2.22	2.74	2.47	3.98	1.35	1.11	1.47	1.62	7.64	4.13
D/O	LA	5.80	5.01	2.63	3.50	4.03	4.09	3.31	6.56	2.91	1.99	3.97	2.23	2.84	4.74	3.68	2.98	5.12	2.76	2.51	2.94	1.30	2.32	2.36	2.49
	NM	7.02	10.45	12.07	6.14	14.43	31.86	6.55	9.26	18.90	2.48	8.45	12.04	5.47	11.90	3.25	4.11	13.18	4.81	8.10	8.08	4.52	5.71	6.27	14.42
	SF	7.12	8.29	7.51	7.80	5.43	7.68	4.01	8.37	6.11	5.03	4.40	4.19	3.52	4.79	4.80	5.93	4.30	7.24	2.56	6.64	5.53	4.33	6.31	4.41
	SH2	3.70	10.83	7.42	4.22	6.12	3.25	2.86	4.62	4.48	3.31	2.89	2.02	3.34	5.53	3.62	2.11	3.58	2.27	3.38	2.42	2.20	2.84	1.43	2.92
	SM	4.03	9.56	6.78	5.39	10.92	6.43	4.39	7.69	8.53	6.69	5.05	6.45	9.31	6.85	7.35	7.91	20.11	14.43	5.73	7.55	4.57	7.08	8.03	13.37
	VA	2.88	2.28	1.82	1.83	4.27	2.12	1.98	1.35	1.96	1.80	2.10	2.11	2.56	4.74	3.40	2.49	3.48	1.69	2.34	2.38	1.96	2.49	1.83	1.78

		Block 5						Block 6						Block 7						Block 8					
		Red			Blue			Red			Blue			Red			Blue			Red			Blue		
Cs	Ps	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq	Sym	Tr	Eq
R/R	DS1																								
	JD																								
	JP																								
	KM																								
	MR																								
	TJ																								
R/O	DS2	2.50	3.36	2.67	7.01	9.00	8.23	2.58	3.93	6.17	5.21	6.32	7.40	4.31	2.57	1.99	5.19	6.70	5.17	2.04	3.63	2.33	4.64	6.23	7.76
	JS																								
	MH	2.57	2.84	1.33	4.18	6.71	4.52	1.92	1.73	1.79	3.24	4.00	4.88	1.96	1.97	2.77	2.29	4.58	4.67	1.95	1.78	1.48	3.16	4.58	4.40
	OR	1.43	0.96	1.53	2.02	2.54	2.74																		
	SJ	2.07	2.08	2.74	2.50	3.61	3.24	1.98	1.59	3.36	1.69	3.99	2.67	1.56	1.50	2.81	2.38	3.59	2.72	1.59	1.29	1.83	2.93	2.10	2.43
	SH1	2.17	1.73	1.60	4.99	4.76	3.41	3.89	3.20	2.21	3.53	3.08	5.99												
R/D	AF																								
	BL	1.31	1.25	2.37	5.77	7.82	7.17	2.03	1.90	2.21	4.40	7.21	10.36	1.88	2.70	3.60	5.35	6.31	6.80	1.93	1.60	3.30	3.07	10.54	9.32
	CP	2.67	1.83	1.75	2.52	8.64	6.40	1.36	1.61	2.08	1.79	3.13	3.62	1.42	1.37	1.66	2.38	4.78	4.31						
	KP	1.59	1.61	1.83	2.96	5.36	11.91	1.84	1.66	1.83	2.60	3.90	3.33												
	MP	5.53	8.48	7.00	7.55	5.15	4.51																		
	RS	1.40	1.53	1.39	2.54	1.63	2.94	1.72	1.26	1.08	2.28	3.14	5.22	1.73	1.67	1.47	3.01	2.72	5.02	1.44	1.14	1.18	2.59	2.79	3.73
D/O	LA	4.41	3.12	2.66	2.90	3.33	5.22	2.55	3.58	3.02	1.52	2.37	2.19	3.94	3.86	2.95	3.01	2.08	2.21	1.73	0.88	1.31	1.09	1.32	1.29
	NM	2.00	1.75	1.44	3.16	1.36	4.75	2.77	3.05	6.63	3.40	10.10	4.02	3.25	5.62	2.25	5.01	3.68	1.93	2.22	3.48	5.14	2.35	3.75	4.53
	SF	4.08	9.05	3.59	5.39	8.42	6.08	2.68	4.36	4.87	4.76	3.98	4.53	2.98	5.40	4.60	4.50	4.53	3.65	1.94	7.05	5.16	3.87	4.47	5.37
	SH2	4.29	9.33	7.90	3.73	6.50	2.68	2.39	3.24	5.93	2.19	2.13	2.09	2.50	6.07	2.63	2.93	3.07	2.20	2.57	2.33	4.32	3.65	2.04	1.60
	SM	4.27	6.21	10.25	6.81	8.04	12.47	4.22	5.66	9.63	3.98	6.81	8.38	5.21	4.46	6.57	5.31	10.82	7.87	3.48	5.41	7.40	4.58	9.18	7.98
	VA	2.01	4.85	2.83	2.46	5.90	2.96	2.26	2.95	2.34	1.81	2.31	1.67	2.38	3.82	1.97	1.54	2.38	1.69	2.09	1.74	2.35	1.92	3.06	2.00

APPENDIX G

Study Four: Trials (Tr) required and errors (Er) made by all participants in the protocol (Pr), mixed (Mx), and no protocol (Np) conditions (Cs) during each stage of match-to-sample training and testing: Phase 1, Reinforcement baseline training (Bt+); Phase 2, Extinction baseline review (Br-); Phase 3, Emergent testing (Emgt). Blank cells indicate that no training or testing occurred during that phase.

		Bt		Br		Emgt1		Emgt2		Br		Bt		Emgt3		Emgt4	
Cs	Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
Pr	AD	300	86	12	0	24	6	24	10	24	1	60	5	24	10	24	8
	CO	60	21	12	0	24	1										
	HZ	288	122	12	0	24	8	24	7	12	0			24	11	24	8
	JC	276	114	12	0	24	4	24	7	36	5	48	3	24	10	24	7
	KS	468	215														
	LC	192	133														
	LD	192	112														
	LB	84	19	24	1	24	0										
	OH	204	108														
	RC	564	204														
	SL	276	63														
	SC	36	19	12	0	24	0										
	TP	84	36	24	1	24	0										
Mx	AK	108	25	12	0	24	0										
	CD	72	19	12	0	24	0										
	CR	276	110	12	0	24	7	24	8	24	1	36	2	24	6	24	9
	DW	252	88	24	1	24	4	24	2	24	1	12	0	24	13	24	9
	DF	216	85	12	0	24	0										
	EJ	396	158	24	1	24	8	24	13	24	7	24	1	24	3	24	0
	JM	48	4	24	1	24	2	24	0								
	JP	168	31	12	0	24	3	24	2					24	0		
	JB	60	8	24	2	24	0										
	LF	276	99	12	0	24	5	24	2			24	2	24	4	24	4
	LB	72	13	12	0	24	0										
	SK	144	32	24	1	24	2	24	0								
	VB	228	63	12	0	24	6	24	3	36	2	48	2	24	2	24	4
Np	AR	48	10	12	0	24	2	24	0								
	DM	228	57	24	1	24	0										
	DC1	72	26	24	1	24	0										
	DC2	36	6	12	0	24	0										
	FP	168	78	12	0	24	1										
	JF	72	10	12	0	24	0										
	JC	204	70	24	5	24	0										
	MF	96	24	12	0	24	0										
	MW	108	25	12	0	24	0										
	PM	36	5	12	0	24	1										
	RS	144	35	12	0	24	2	24	0								
	RG	48	7	12	0	24	2	24	0								
	TS	24	4	12	0	24	0										

APPENDIX H

Study Four: Session durations (in minutes, rounded to nearest minute) for all participants in the protocol (Pr), mixed (Mx), and no protocol (Np) conditions (Cs). ^{T} indicates termination of the experiment by experimenter, ^{R} indicates termination of the experiment at request of participant.

Cs	Ps	Duration
Pr	AD	70
	CO	37
	HZ	88
	JC	60
	KS	60^{T}
	LC	50^{R}
	LD	60^{T}
	LB	36
	OH	46^{R}
	RC	60^{T}
	SL	34^{R}
	SC	17
	TP	28
Mx	AK	30
	CD	19
	CR	58
	DW	49
	DF	33
	EJ	60
	JM	21
	JP	35
	JB	22
	LF	59
	LB	13
	SK	28
	VB	57
Np	AR	17
	DM	45
	DC1	16
	DC2	8
	FP	37
	JF	17
	JC	42
	MF	21
	MW	24
	PM	12
	RS	33
	RG	18
	TS	10

APPENDIX I

Study Five, Experiment One: Trials (Tr) required and errors (Er) made by all participants (Ps) during each stage of match-to-sample training and testing: Phase 1, Reinforcement training (Rt); Phase 2, Initial emergent testing (1Em); Phase 3, Secondary emergent testing (2Em); Phase 4, Generalisation testing (Gn). Blank cells indicate that no training or testing occurred during that phase. Data rejected from analysis (participants CY and KY) are presented separately.

	Rt1		1Em1		1Em2		1Em3		Rt2		1Em4		1Em5		1Em6		Rt3	
Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
CB	1	0	12	3	12	1	12	0										
KS	4	0	12	4	12	0												
JM	12	0	12	0														
CY	12	0	12	4	12	4	12	3	12	0	12	4	12	4	12	8	12	0
KY	12	0	12	8	12	7	12	8	12	0	12	11	12	11	12	11	12	0

	1Em7		1Em8		1Em9		Rt4		1Em10		1Em11		1Em12		2Em1		Gn1	
Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
CB															36	0	18	0
KS															36	0	18	0
JM															36	0	18	0
CY	12	5	12	3	12	5	12	0	12	5	12	5	12	3				
KY	12	10	12	10	12	10	12	0	12	4	12	4	12	5				

APPENDIX J

Study Five, Experiment Two: Trials (Tr) required and errors (Er) made by participants (Ps) AM, HR, RC, RM, AF, and IN during each stage of match-to-sample training and testing: Phase 1, Initial emergent testing (1Em); Phase 2, Reinforcement training (Rt); Phase 3, Secondary emergent testing (2Em); Phase 4, Generalisation testing (Gn). Blank cells indicate that no training or testing occurred during that phase.

	1Em1		1Em2		1Em3		Rt1		1Em4		1Em5		1Em6		Rt2		1Em7	
Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
AM	12	4	12	0														
HR	12	2	12	0														
RC	12	1	12	1	12	0												
RM	12	10	12	9	12	8	4	0	12	9	12	10	12	9	4	0	12	8
AF	12	9	12	8	12	9	12	0	12	6	12	0						
IN	12	3	12	1	12	2	12	0	12	1	12	0						

	1Em8		1Em9		2Em1		2Em2		1Em10		2Em3		2Em4		Gn1		Gn2	
Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
AM					36	2	36	0							18	0		
HR					36	0									18	0		
RC					36	0									18	0		
RM	12	5	12	0	36	4	36	2	12	0	36	4	36	1	18	1	18	0
AF					36	0									18	0		
IN					36	0									18	0		

APPENDIX K

Study Five, Experiment Two: Trials (Tr) required and errors (Er) made by participants (Ps) AC, PS, MM, PD, RO, and WK during each stage of match-to-sample training and testing: Phase 1, Initial emergent testing (1Em); Phase 2, Reinforcement training (Rt); Phase 3, Secondary emergent testing (2Em); Phase 4, Generalisation testing (Gn). Blank cells indicate that no training or testing occurred during that phase.

	1Em1		1Em2		1Em3		1Em4		1Em5		1Em6		1Em7		1Em8		1Em9		1Em10		1Em11		1Em12		Rt1	
Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
AC	12	4	12	2	12	1	12	0																		
PS	12	11	12	10	12	10	12	11	12	10	12	10	12	11	12	11	12	11	12	11	12	12	12	11	1	0
MM	12	7	12	9	12	7	12	5	12	5	12	8	12	10	12	7	12	10	12	8	12	9	12	9	4	0
PD	12	1	12	2	12	1	12	1	12	2	12	1	12	1	12	0										
RO	12	9	12	9	12	10	12	8	12	9	12	9	12	6	12	5	12	5	12	5	12	1	12	2	12	0
WK	12	0																								

	1Em13		1Em14		1Em15		Rt2		1Em16		1Em17		2Em1		2Em2		1Em18		2Em3		2Em4		Gn1		Gn2	
Ps	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er	Tr	Er
AC													36	1	36	0							18	1	18	0
PS	12	11	12	8	12	11	1	0	12	0			36	0									18	0		
MM	12	10	12	9	12	8	4	0	12	2	12	0	36	0									18	0		
PD													36	2	36	3	12	0	36	3	36	8	18	2	18	1
RO	12	2	12	0									36	0									18	1	18	2
WK													36	1	36	1	12	0	36	0			18	0		

REFERENCES

Adams, B. J., Fields, L., & Verhave, T. (1993a). Effects of test order on intersubject variability during equivalence class formation. *The Psychological Record,* **43,** 133-152.

Adams, B. J., Fields, L., & Verhave, T. (1993b). Formation of generalized equivalence classes. *The Psychological Record,* **43,** 553-566.

Annett, J. M., & Leslie, J. C. (1995). Stimulus equivalence classes involving olfactory stimuli. *The Psychological Record,* **45,** 439-450.

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record,* **47,** 309-320.

Augustson, E. M., & Dougher, M. J. (1997). The transfer of avoidance evoking functions through stimulus equivalence classes. *Journal of Behavior Therapy and Experimental Psychiatry,* **28,** 181-191.

Austin, J., & Delaney, P. F. (1998). Protocol analysis as a tool for behaviour analysis. *The Analysis of Verbal Behavior,* **15,** 41-56.

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record,* **44,** 91-124.

Barnes, D. (1996). Naming as a technical term: Sacrificing behavior analysis at the altar of popularity [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 264-267.

Barnes, D., Browne, M., Smeets, P., & Roche, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three- to six-year-old children. *The Psychological Record,* **45,** 405-430.

Barnes, D., & Hampson, P. J. (1993). Stimulus equivalence and connectionism: Implications for behavior analysis and cognitive science. *The Psychological Record,* **43,** 617-638.

Barnes, D., & Holmes, Y. (1991). Radical behaviorism, stimulus equivalence, and human cognition. *The Psychological Record,* **41,** 19-31.

Barnes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior,* **59,** 61-81.

Barnes, D., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired and hearing impaired children. *The Analysis of Verbal Behavior,* **8,** 19-30.

Barnes, D., & Roche, B. (1996a). Relation frame theory and the experimental analysis of human sexuality. *Applied and Preventive Psychology,* **6,** 117-135.

Barnes, D., & Roche, B. (1996b). Relational frame theory and stimulus equivalence are fundamentally different: A reply to Saunders' commentary. *The Psychological Record,* **46,** 489-507.

Barnes, D., & Roche, B. (1997). A behavior-analytic approach to behavioral reflexivity. *The Psychological Record,* **47,** 543-572.

Belanich, J., & Fields, L. (1999). Tactual equivalence class formation and tactual-to-visual cross-modal transfer. *The Psychological Record,* **49,** 75-91.

Bentall, R. P., Dickins, D. W. and Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology,* **46B,** 187-214.

Bentall, R. P., Jones, R. M., & Dickins, D. W. (1999). Errors and response latencies as a function of nodal distance in 5-member equivalence classes. *The Psychological Record,* **49,** 93-115.

Bentall, R. P., Lowe, C. F., & Beasty, A. (1985). The role of verbal behavior in human learning: II. Developmental differences. *Journal of the Experimental Analysis of Behavior,* **43,** 165-181.

Bonta, J. L., & Waters, R. G. (1981). Use of manual signs in delayed matching-to-sample with developmentally disordered, speech deficient children. *Behavior Research of Severe Developmental Disabilities,* **2,** 51-66.

Broadbent, D. (1986). Protocol analysis: Verbal reports as data. *Quarterly Journal of Experiment Psychology,* **38A,** 524-525.

Brown, A. K., Brown, J. L., & Poulson, C. L. (1995). Generalization of children's identity match-to-sample performances to novel stimuli. *The Psychological Record,* **45,** 29-43.

Brown, L., Ed. (1993). *The New Shorter English Dictionary.* Oxford: Oxford University Press.

Buffington, D. M., Fields, L., & Adams, B. J. (1997). Enhancing equivalence class formation by pre-training of other equivalence classes. *The Psychological Record,* **47,** 69-96.

Bush, K. M. (1993). Stimulus equivalence and cross-modal transfer. *The Psychological Record,* **43,** 567-584.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior,* **51,** 29-45.

Carr, D. (1997). Stimulus equivalence, naming and contextual control: Studies with language disabled autistic adolescents. Unpublished doctoral dissertation, University of Wales College, Cardiff.

Carr, D., & Blackman, D. E. (1996). Equivalence relations, naming and generalised symmetry [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 245-247.

Carrigan, P. F. Jr., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior,* **58,** 183-204.

Chase, P. N. (1996). The origins of naming: A critique of self-listening [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 293-296.

Chiesa, M. (1992). Radical behaviorism and scientific frameworks: From mechanistic to relational accounts. *American Psychologist,* **47,** 1287-1299.

Chiesa, M. (1994). Radical Behaviorism: The philosophy and the science. Boston: Authors Cooperative.

Clayton, M. C., & Hayes, L. J. (1999). Conceptual differences in the analysis of stimulus equivalence. *The Psychological Record,* **49,** 145-161.

Commons, M. L., & Rodriguez, J. A. (1993). The development of hierarchically complex equivalence classes. *The Psychological Record,* **43,** 667-695.

Constantine, B., & Sidman, M. (1975). Role of naming in delayed matching to sample. *American Journal of Mental Deficiency,* **79,** 680-689.

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis,* **25,** 461-475.

Critchfield, T. S., & Perone, M. (1990). Verbal self-reports of delayed matching-to-sample by humans. *Journal of the Experimental Analysis of Behavior,* **53,** 321-344.

Critchfield, T. S., Tucker, J. A., & Vuchinich, R. E. (1998). Self-report methods. In K. A. Lattal and M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 435-470). NY, Plenum Press.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science,* **5,** 241-244.

Cullinan, V. A., Barnes, D., & Smeets, P. M. (1998). A precursor to the relational evaluation procedure: Analyzing stimulus equivalence. *The Psychological Record,* **48,** 121-145.

D'Amato, M. R., Salmon, D. P., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Journal of the Experimental Analysis of Behavior,* **44,** 35-47.

DeGrandpre, R. J., & Bickel, W. K. (1993), Stimulus control and drug dependence. *The Psychological Record,* **43,** 651-666.

DeGrandpre, R. J., Bickel, W. K., & Higgins, S. T. (1992). Emergent equivalence relations between interoceptive (drug) and exteroceptive (visual) stimuli. *Journal of the Experimental Analysis of Behavior,* **58,** 9-18.

de Rose, J. C. (1996). Naming, meaning, and verbal operants [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 274-276.

de Rose, J. C., de Souza, D. G., & Hanna, E.S. (1996). Teaching reading and spelling: Exclusion and stimulus equivalence. *Journal of Applied Behavior Analysis,* **29,** 451-470.

de Rose, J. C., McIlvane, W. J., Dube, W. V., & Stoddard, L. T. (1988). Stimulus class formation and functional equivalence in moderately retarded individuals' conditional discrimination. *Behavioural Processes,* **17,** 167-175.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior,* **46,** 243-257.

Dickins, D. W., Bentall, R. P., & Smith, A. B. (1993). The role of individual stimulus names in the emergence of equivalence relations: The effects of interpolated paired-associates training of discordant associations between names. *The Psychological Record,* **43,** 713-724.

Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior,* **27,** 433-442.

Dixon, M., & Spradlin, J. (1976). Establishing stimulus equivalences among retarded adolescents. *Journal of Experimental Child Psychology,* **21,** 144-164.

Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior,* **62,** 331-351.

Duarte, A. M., Eikeseth, S., Rosales-Ruiz, J., & Baer, D. M. (1998). The effects of a can't answer response option and instructions on stimulus equivalence. *The Psychological Record,* **48,** 631-646.

Dube, W. V., Green, G., & Serna, R. W. (1993). Auditory successive conditional discrimination and auditory stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior,* **59,** 103-114.

Dube, W. V., & Hiris, E. (1996). Match to sample program (version 10.18) [Computer software]. Waltham, MA: E. K. Shriver Center.

Dube, W. V., Iennaco, F. M., Rocco, F. J., Klederas, J. B., & McIlvane, W. J. (1992). Microcomputer-based programmed instruction in identity matching to sample for persons with severe disabilities. *Journal of Behavioral Education,* **2,** 29-51.

Dube, W. V., & McIlvane, W. J. (1997). Reinforcer frequency and restricted stimulus control. *Journal of the Experimental Analysis of Behavior,* **68,** 303-316.

Dube, W. V., McIlvane, W. J., Callahan, T. D., & Stoddard, L. T. (1993). The search for stimulus equivalence in nonverbal organisms. *The Psychological Record,* **43,** 761-778.

Dube, W. V., McIlvane, W. J., Mackay, H. A., & Stoddard, L. T. (1987). Stimulus class membership established via stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior,* **47,** 159-175.

Dube, W. V., McIlvane, W. J., Maguire, R. W., Mackay, H. A., & Stoddard, L. T. (1989). Stimulus class formation and stimulus-reinforcer relations. *Journal of the Experimental Analysis of Behavior,* **51,** 65-76.

Dugdale, N., & Lowe, C.F. (1990). Naming and stimulus equivalence. In D. E. Blackman and H. Lejeune (Eds.), *Behaviour analysis in theory and practice:*

Contributions and controversies (pp. 115-138). Hove, England: Lawrence Erlbaum Associates.

Duncan, C. P. (1985). Protocol analysis: Verbal reports as data. *American Journal of Psychology,* **98.** 641-642.

Dymond, S., & Barnes, D. (1994). A transformation of self-discrimination response functions through equivalence relations. *Journal of the Experimental Analysis of Behavior,* **62,** 251-267.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior,* **64,** 163-184.

Dymond, S., & Barnes, D. (1996). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness and opposition. *The Psychological Record,* **46,** 271-300.

Dymond, S., & Barnes, D. (1997), The effects of prior equivalence testing and verbal instructions on derived self-discrimination transfer: A follow-up study. *The Psychological Record,* **48,** 147-170.

Eikeseth, S., Rosales-Ruiz, J., Duarte, A., & Baer, D. M. (1997). The quick development of equivalence classes in a paper-and-pencil format through written instructions. *The Psychological Record,* **47,** 275-284.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior,* **58,** 123-133.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review,* **87,** 215-249.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA. MIT Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity,* **5,** 178-186.

Fields, L. (1996). The evidence for naming as a cause or facilitator of equivalence class formation [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 279-282.

Fields, L., Adams, B. J., Buffington, D. M., Yang, W., & Verhave, T. (1996). Response transfer between stimuli in generalized equivalence classes: A model for the establishment of natural kind and fuzzy superordinate categories. *The Psychological Record,* **46,** 665-684.

Fields, L., Adams, B. J., & Verhave, T. (1993). The effects of equivalence class structure on test performances. *The Psychological Record,* **43,** 697-712.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior,* **53,** 345-358.

Fields, L., Adams, B. J., Verhave, T., & Newman, S. (1993). Are stimuli in equivalence classes equally related to each other? *The Psychological Record,* **43,** 85-105.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalisation gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior,* **68,** 67-91.

Fields, L., Reeve, K. F., Adams, B. J., Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior,* **55,** 305-312.

Fields, L., Reeve, K. F., Rosen, D., Varelas, A., Adams, B. J., Belanich, J., & Hobbie, S. A. (1997). Using the simultaneous protocol to study equivalence class formation: The facilitating effects of nodal number and size of previously established equivalence classes. *Journal of the Experimental Analysis of Behavior,* **67,** 367-389.

Fields, L., Reeve, K. F., Varelas, A., Rosen, D., & Belanich, J. (1997). Equivalence class formation using the stimulus-pairing and yes-no responding. *The Psychological Record,* **47,** 661-666.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior,* **48,** 317-322.

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior,* **42,** 143-157.

Gast, D. L., VanBiervliet, A., & Spradlin, J. E. (1979). Teaching number-word equivalences: A study of transfer. *American Journal of Mental Deficiency,* **83,** 524-527.

Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal of the Experimental Analysis of Behavior,* **51,** 369-378.

Gisiner, R., & Schusterman, R. J. (1992). Sequence, syntax, and semantics: Responses of a language-trained sea lion (*Zalophus californianus*) to novel sign combinations. *Journal of Comparative Psychology,* **106,** 78-91.

Glat, R., Gould, K., Stoddard, L. T., & Sidman, M. (1994). A note on transfer of stimulus control in the delayed-cue procedure: Facilitation by an overt differential response. *Journal of Applied Behavior Analysis,* **27,** 699-704.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hove, England: Lawrence Erlbaum Associates.

Green, G. (1990). Differences in development of visual and auditory-visual equivalence relations. *American Journal on Mental Retardation,* **95,** 260-270.

Green, G., Mackay, H. A., McIlvane, W. J., Saunders, R. R., & Soraci, S. A. Jnr. (1990). Perspectives on relational learning in mental retardation. *American Journal on Mental Retardation,* **95,** 304-315.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal and M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229-262). NY, Plenum Press.

Green, G., Sigurdardottir, Z. G., & Saunders, R. R. (1991). The role of instructions in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior,* **55,** 287-304.

Green, G., Stromer, R., & Mackay, H. A. (1993). Relational learning in stimulus sequences. *The Psychological Record,* **43,** 599-616.

Greenway, D. E., Dougher, M. J., & Wulfert, E. (1996). Transfer of consequential functions via stimulus equivalence: Generalization to different testing contexts. *The Psychological Record,* **46**, 131-143.

Grey, I. M., & Barnes, D. (1996). Stimulus equivalence and attitudes. *The Psychological Record,* **46,** 243-270.

Hall, G. A., & Chase, P. N. (1991). The relationship between stimulus equivalence and verbal behaviour. *The Analysis of Verbal behaviour,* **9,** 107-119.

Haring, T. G., Breen, C. G., & Laitinen, R. E. (1989). Stimulus class formation and concept learning: Establishment of within- and between-set generalization and transitive relationships via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior,* **52,** 13-25.

Harnad, S. (1996). Experimental analysis of naming behavior cannot explain naming capacity [commentary]. *Journal of the Experimental Analysis of behavior,* **65,** 262-264.

Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior,* **54,** 225-237.

Hayes, L. J., Thompson, S., & Hayes, S. C. (1989). Stimulus equivalence and rule following. *Journal of the Experimental Analysis of Behavior,* **52,** 275-291.

Hayes, L. J., Tilley, K. J., & Hayes, S. C. (1988). Extending equivalence class membership to gustatory stimuli. *The Psychological Record,* **38,** 473-482.

Hayes, S. C. (1986). The case of the silent dog--verbal reports and the analysis of rules: A review of Ericsson and Simon's protocol analysis: Verbal reports as data. (1986). *Journal of the Experimental Analysis of Behavior,* **45,** 351-363.

Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **51,** 385-392.

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behaviour* (pp. 19-40). Reno, NV: Context Press.

Hayes, S. C. (1992). Verbal relations, time, and suicide. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 109-118). Reno, NV: Context Press.

Hayes, S. C. (1994). Relational frame theory: A functional approach to verbal events. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition* (pp. 9-30). Reno, NV: Context Press.

Hayes, S. C. (1996). Developing a theory of derived stimulus relations [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 309-311.

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior,* **68,** 235-270.

Hayes, S. C., & Gifford, E. V. (1997). The trouble with language: Experiential avoidance, rules, and the nature of verbal events. *Psychological Science,* **8,** 170-173.

Hayes, S. C., Gifford, E. V., & Hayes, G. J. (1998). Moral behavior and the development of verbal regulation. *The Behavior Analyst,* **21,** 253-279.

Hayes, S. C., & Hayes, L. J. (1989). The verbal action of the listener as a basis for rule-governance. In S. C. Hayes (Ed.), *Rule-governed behavior: Cognition, contingencies, and instructional control* (pp. 153-190). New York: Plenum.

Hayes, S. C., & Hayes, L. J. (1992). Verbal relations and the evolution of behavior analysis. *American Psychologist,* **47,** 1383-1395.

Hayes, S. C., Kohlenberg, B. S., & Hayes, L. C. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior,* **56,** 119-137.

Hayes, S. C., White, D., & Bissett, R. T. (1998). Protocol analysis and the "silent dog" method of analyzing the impact of self-generated rules. *The Analysis of Verbal Behavior,* **15,** 57-63.

Hayes, S. C., & Wilson, K. G. (1996). Criticisms of relational frame theory: Implications for a behaviour analytic account of derived stimulus relations. *The Psychological Record,* **46,** 221-236.

Holth, P., & Arntzen, E. (1998). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record,* **48,** 81-110.

Horne, P. J., & Lowe, C.F. (1993). Determinants of human performance on concurrent schedules. *Journal of the Experimental Analysis of Behavior,* **59,** 29-60.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of behavior,* **65,** 185-243.

Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behaviour. *Journal of the Experimental Analysis of Behavior,* **68,** 271-296.

Innis, A., Lane, S. D., Miller, E. R., & Critchfield, T. S. (1998). Stimulus equivalence: Effects of a default-response option on emergence of untrained stimulus relations. *Journal of the Experimental Analysis of Behavior,* **70,** 87-102.

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.* New York: McGraw-Hill.

Jenkins, J. J., & 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 Social Research and Child Development,* **29,** No. 1, 141-169.

Johnson, C. and Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior,* **59,** 333-347.

Joyce, B., & Joyce, J. (1990, May). *Stimulus equivalence: An approach for training children with TBI.* Poster presented at the meeting of the Association for Behavior Analysis, Nashville, TN.

Joyce, B. G., & Wolking, W. D. (1989). Stimulus equivalence: An approach for teaching beginning reading skills to young children. *Education and Treatment of Children,* **12,** 109-122.

Kamin, L. H. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and Aversive behavior*(pp. 279-296). New York: Appleton-Century Crofts.

Kennedy, C. H. (1991). Equivalence class formation influenced by the number of nodes separating stimuli. *Behavioural Processes,* **24,** 219-245.

Kennedy, C. H., Itkonen, T., & Lindquist, K. (1994). Nodality effects during equivalence class formation: An extension to sight-word reading and concept development. *Journal of Applied Behavior Analysis,* **27,** 673-683.

Kennedy, C. H., & Laitinen, R. (1988). Second-order conditional control of symmetric and transitivity stimulus relations: The influence of order effects. *The Psychological Record,* **38,** 437-446.

Kennedy, C. H., & Serna, L. A. (1995). Emergent relational responding based upon quantity and equivalence. *The Psychological Record,* **45,** 241-260.

Kirby, K. C., & Bickel, W. K. (1988). Toward an explicit analysis of generalization: A stimulus control interpretation. *The Behavior Analyst,* **11,** 115-129.

Kohlenberg, B. S., Hayes, S. C., & Hayes, L. J. (1991). The transfer of contextual control over equivalence classes through equivalence classes: A possible model of social stereotyping. *Journal of the Experimental Analysis of Behavior,* **56,** 505-518.

Kuno, H., Kitadate, T., & Iwamoto, T. (1994). Formation of transitivity in conditional matching to sample by pigeons. *Journal of the Experimental Analysis of Behavior,* **62,** 399-408.

Laffal, J. (1985). Protocol analysis: Verbal reports as data. *Journal of Nervous and Mental Disease,* **173,** 703-704.

Lamarre, J., & Holland, J. G. (1985). The functional independence on mands and tacts. *Journal of the Experimental Analysis of Behavior,* **43,** 5-19.

Lane, S. D., Clow, J. K., Innis, A., & Critchfield, T. S. (1998). Generalization of cross-modal stimulus equivalence classes: Operant processes as components in category formation. *Journal of the Experimental Analysis of Behavior,* **70,** 267-279.

Lane, S. D., & Critchfield, T. S. (1996). Verbal self-reports of equivalence relations in a stimulus equivalence procedure. *Journal of the Experimental Analysis of Behavior,* **65,** 355-374.

Lazar, R. (1977). Extending sequence-class membership with matching to sample. *Journal of the Experimental Analysis of Behavior,* **27,** 381-392.

Lazar, R., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in children. *Journal of the Experimental Analysis of Behavior,* **41,** 251-266.

Lazar, R. M., & Kotlarchyk, B. J. (1986). Second-order control of sequence-class equivalences in children. *Behavioural processes,* **13,** 205-215.

Lazar, R. M., & Scarisbrick, D. (1993). Alexia without agraphia: A functional assessment of behavior in focal neurologic disease. *The Psychological Record,* **43,** 639-650.

Leader, G., Barnes, D., & Smeets, P. M. (1996). Establishing equivalence relations using a respondent-type training procedure. *The Psychological Record,* **46,** 685-706.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longditudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology,* **56,** 201-239.

Lipkens, R., Kop, P. F. M., & 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.

Lowe, C. F. (1987). The role of language in the emergence of equivalence relations. *Bulletin of the British Psychological Society,* **40,** A49.

Lowe, C. F., & Beasty, A. (1987). Language and the emergence of equivalence relations: A developmental study. *Bulletin of the British Psychological Society,* **40,** A49.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behaviour. *Journal of the Experimental Analysis of Behavior,* **65,** 315-340.

Lowenkron, B. (1988). Generalization of delayed identity matching in retarded children. *Journal of the Experimental Analysis of Behavior,* **50,** 163-172.

Lowenkron, B. (1996). Joint control and word-object bidirectionality [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 252-255.

Lowenkron, B. (1997). The role of joint control in the development of naming [commentary]. *Journal of the Experimental Analysis of Behavior,* **68,** 244-247.

Lowenkron, B., & Colvin, V. (1995). Generalized instructional control and the production of broadly applicable relational responding. *The Analysis of Verbal behavior,* **12,** 13-29.

Lynch, D. C., & Cuvo, A. J. (1995). Stimulus equivalence instruction of fraction-decimal relations. *Journal of Applied Behavior Analysis,* **28,** 115-126.

Lynch, D. C., & Green, G. (1991). Development and crossmodal transfer of contextual control of emergent stimulus relations. *Journal of the Experimental Analysis of Behavior,* **56,** 139-154.

MacDonald, R. P. F., Dixon, L. S., & LeBlanc, J. M. (1986). Stimulus class formation following observational learning. *Analysis and Intervention in Developmental Disabilities,* **6,** 73-87.

Mack, R. L. (1985). Verbal protocol analysis comes of age. *Contemporary Psychology,* **30,** 373-375.

Mackay, H. A. (1985). Stimulus equivalence in rudimentary reading and spelling. *Analysis and Intervention in Developmental Disabilities,* **5,** 373-387.

Mackay, H. A. (1991). Stimulus equivalence: Implications for the development of adaptive behavior. In R. Remington (Ed.), *The Challenge of severe mental handicap: An applied behaviour analytic approach.* (pp. 235-259). Chichester: Wiley.

Mackay, H. A., & Ratti, C. A. (1990). Position-numeral equivalences and delayed position recognition span. *American Journal on Mental Retardation,* **95,** 271-282.

Mackay, H. A., & Sidman, M. (1984). Teaching new behaviour via equivalence relations. In P. H. Brooks, R. Sperber, & C. Macauley (Eds.), *Learning and cognition in the mentally retarded* (pp. 493-513). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mackintosh, J. H. (1977). Stimulus control: Attentional factors. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior.* New Jersey: Prentice-Hall.

Magliano, J. P. (1996). Verbal protocol analysis revisited. *Applied Cognitive Psychology,* **10,** 273-274.

Maguire, R. W., Stromer, R., & Mackay, H. A. (1995). Delayed matching to complex samples and the formation of stimulus classes in children. *Psychological Reports,* **77,** 1059-1076.

Maguire, R. W., Stromer, R., Mackay, H. A., & Demis, C. A. (1994). Matching to complex samples and stimulus class formation in adults with autism and young children. *Journal of Autism and Developmental Disorders,* **24,** 753-772.

Manabe, K., Kawashima, T., & Staddon, J. E. R. (1995). Differential vocalization in budgerigars: Towards an experimental analysis of naming. *Journal of the Experimental Analysis of Behavior,* **63,** 111-126.

Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural processes,* **32,** 29-46.

Markham, M. R., & Dougher, M. J. (1993). Compound stimuli in emergent stimulus relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **60,** 529-542.

Maydak, M., Stromer, R., Mackay, H. A., & Stoddard, L. T. (1995). Stimulus classes in matching to sample and sequence production: The emergence of numeric relations. *Research in Developmental Disabilities,* **16,** 179-204.

McDonagh, E. C., McIlvane, W. J., & Stoddard, L. T. (1984). Teaching coin equivalences visual matching to sample. *Applied Research in Mental Retardation,* **5,** 177-197.

McIlvane, W. J., & Dube, W. V. (1996). Naming as a facilitator of discrimination [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 267-272.

McIlvane, W. J., Dube, W. V., Klederas, J. B., Ienacco, F. M., & Stoddard, L. T. (1990). Teaching relational discriminations to individuals with mental retardation: Some problems and possible solutions. *American Journal on Mental Retardation,* **95,** 283-296.

McIntire, K. D., Cleary, J., & Thompson, T. (1987). Conditional relations by monkeys: Reflexivity, symmetry, and transitivity. *Journal of the Experimental Analysis of Behavior,* **47,** 279-285.

Meehan, E. F. (1995). Emergence by exclusion. *The Psychological Record,* **45,** 133-154.

Meehan, E. F., & Fields, L. (1995). Contextual control of new equivalence classes. *The Psychological Record,* **45,** 165-182.

Michael, J. (1996). Separate repertoires or naming? [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 296-298.

Nisbett. R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports of mental processes. *Psychological Review,* **87,** 231-259.

O'Leary, C. A., & Bush, K. M. (1996). Stimulus equivalence in the tactile modality. *The Psychological Record,* **46,** 509-517.

O'Mara, H. (1991). Quantitative and methodological aspects of stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **55,** 125-132.

Osbourne, J. G., & Gatch, M. B. (1989). Stimulus equivalence and receptive reading by hearing-impaired preschool children. *Language, Speech, and Hearing Services in Schools,* **20,** 63-75.

Payne, J. W. (1994). Thinking-aloud: Insights into information processing. *Psychological Science,* **5,** 241-248.

Payne, J. W., Braunstein, M. L., & Carroll, J. S. (1978). Exploring predecisional behavior: An alternative approach to decision research. *Organizational Behavior and Human Performance,* **22,** 17-44.

Peláez-Nogueras, M. (1996). Thought without naming [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 299-301.

Peoples, M., Tierney, K. J., Bracken, M., & McKay, C. (1998). Prior learning and equivalence class formation. *The Psychological Record,* **48,** 111-120.

Pepperberg, I. M. (1990). Cognition in an African Gray parrot (Psittacus erithacus): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology,* **104,** 41-52.

Perone, M. (1988). Laboratory lore and research practices in the experimental analysis of human behavior: Use and abuse of subjects' verbal reports. *The Behavior Analyst,* **11,** 71-75.

Pilgrim, C. (1996). Can the naming hypothesis be falsified? [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 284-286.

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior,* **63,** 239-254.

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior,* **63,** 225-238.

Plaud, J. J. (1995). The formation of stimulus equivalences: Fear-relevant versus fear-irrelevant stimulus classes. *The Psychological Record,* **45,** 207-222.

Plaud, J. J., Gaither, G. A., Franklin, M., Weller, L. A., & Barth, J. (1998). The effects of sexually explicit words on the formation of stimulus equivalence classes. *The Psychological Record,* **48,** 63-79.

Plaud, J. J., Gaither, G. A., Weller, L. A., Bigwood, S. J., Barth, J., & von Duvillard, S. P. (1998). Rational-emotive behavior therapy and the formation of stimulus equivalence classes. *Journal of Clinical Psychology,* **54,** 597-610.

Potter, B., Huber, S., & Michael, J. (1997). The role of mediating verbal behavior in selection-based responding. *The Analysis of Verbal Behavior,* **14,** 41-56.

Pouthas, V., Droit, S., Jacquet, A.-Y., & Wearden, J. H. (1990). Temporal differentiation of response duration in children of different ages: Developmental changes in relations between verbal and nonverbal behavior. *Journal of the Experimental Analysis of Behavior,* **53,** 21-31.

Ransdell, S. (1995). Generating thinking-aloud protocols: Impact on the narrative writing of college students. *American Journal of Psychology,* **108,** 89-98.

Reese, E. P., Howard, J., & Reese, T. W. (1978). *Human Operant Behavior: Analysis and Application.* Iowa: WCB.

Rehfeldt, R. A., Dixon, M. R., Hayes, L. J., & Steele, A. (1998). Stimulus equivalence and the blocking effect. *The Psychological Record,* **48,** 647-664.

Rehfeldt, R. A., & Hayes, L. J. (1998a). The operant-respondent distinction revisited: Toward an understanding of stimulus equivalence. *The Psychological Record,* **48,** 187-210.

Rehfeldt, R. A., & Hayes, L. J. (1998b). Untrained temporal differentiation and equivalence class formation. *The Psychological Record,* **48,** 481-510.

Remington, B. (1996). The evolution of naming--just so! [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 243-245.

Roche, B., & Barnes, D. (1996). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record,* **46,** 451-475.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior,* **67,** 275-301.

Roche, B., & Barnes, D. (1998). The experimental analysis of human sexual arousal: Some recent developments. *The Behavior Analyst,* **21,** 37-52.

Roche, B., Barnes, D., & Smeets, P. M. (1997). Incongruous stimulus pairing and conditional discrimination training: Effects on relational responding. *Journal of the Experimental Analysis of Behavior,* **68,** 143-160.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition,* **17,** 759-769.

Saunders, K. J. (1989). Naming in conditional discrimination and stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **51,** 379-384.

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record,* **43,** 725-744.

Saunders, K. J., & Spradlin, J. E. (1989). Conditional discrimination in mentally retarded adults: The effect of training the component simple discriminations. *Journal of the Experimental Analysis of Behavior,* **52,** 1-12.

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded adults: The development of generalized skills. *Journal of the Experimental Analysis of Behavior,* **54,** 239-250.

Saunders, K. J., & Spradlin, J. E. (1993). Conditional discrimination in mentally retarded subjects: Programming acquisition and learning set. *Journal of the Experimental Analysis of Behavior,* **60,** 571-585.

Saunders, K. J., & Spradlin, J. E. (1996). Naming and equivalence relations [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 304-308.

Saunders, K. J., & Williams, D. C. (1998). Do parakeets exhibit derived stimulus control? Some thoughts on experimental control procedures. *Journal of the Experimental Analysis of Behavior,* **70,** 321-324.

Saunders, R. R. (1996). From review to commentary on Roche and Barnes: Toward a better understanding of equivalence in the context of relational frame theory. *The Psychological Record,* **46,** 477-487.

Saunders, R. R., Drake, K. M., & Spradlin, J. E. (1999). Equivalence class establishment, expansion, and modification in preschool children. *Journal of the Experimental Analysis of Behavior,* **71,** 195-214.

Saunders, R. R., & Green, G. (1992). The nonequivalence of behavioral and mathematical equivalence. *Journal of the Experimental Analysis of Behavior,* **57,** 227-241.

Saunders, R. R., & Green, G. (1996). Naming is not (necessary for) stimulus equivalence [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 312-314.

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior,* **72,** 117-137.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior,* **50,** 145-162.

Saunders, R. R., Saunders, K. J., & Spradlin, J. E. (1990). Long-term stability of equivalence relations in the absence of practice or training. *American Journal on Mental Retardation,* **95,** 297-303.

Saunders, R. R., & Sherman, J. A. (1986). Analysis of the "discrimination-failure hypothesis" in generalized matching and mismatching behavior. *Analysis and Intervention in Developmental Disabilities,* **6,** 89-107.

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior,* **49,** 95-115.

Schenk, J. J. (1993). Emergent conditional discrimination in children: Matching to compound stimuli. *The Quarterly Journal of Experimental Psychology,* **46B,** 345-365.

Schenk, J. J. (1995). Complex stimuli in nonreinforced simple discrimination tasks: Emergent simple and conditional discriminations. *The Psychological Record,* **45,** 477-494.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record,* **43,** 823-839.

Schusterman, R. J., & Kastak, D. (1998). Functional equivalence in a California seal lion: Relevance to animal social and communicative interactions. *Animal Behaviour,* **55,** 1087-1095.

Schusterman, R. J., Kastak, D., & Reichmuth, C. J. (1997). What's in a name? Equivalence by any other name would smell as sweet [commentary]. *Journal of the Experimental Analysis of Behavior,* **68,** 252-257.

Schusterman R. J., & Krieger, K. (1984). California sea lions are capable of semantic comprehension. *The Psychological Record,* **34,** 3-23.

Shimoff, E. (1984). Post-session questionnaires. *Experimental Analysis of Human Behavior Bulletin,* **2,** 1.

Shimoff, E. (1986). Post-session verbal reports and the experimental analysis of behavior. *The Analysis of Verbal Behavior,* **4,** 19-22.

Short, E. J., Evans, S. W., Friebert, S. E., & Schatschneider, C. W. (1991). Thinking aloud during problem solving: Facilitation effects. *Learning and Individual Differences,* **3,** 109-122.

Sidman, M. (1960/1988). *Tactics of Scientific Research: Evaluating Experimental Data in Psychology.* Basic Books. Boston: Authors Cooperative. (reprinted).

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. 2. Education and Training* (pp. 353-360). Baltimore, MD: University Park Press.

Sidman. M. (1981). Remarks. *Behaviorism,* **9,** 127-129.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson and M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213-245). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis,* **22,** 11-18.

Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman and H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93-114). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sidman, M. (1992). Adventitious control by the location of comparison stimuli in conditional discriminations. *Journal of the Experimental Analysis of Behavior,* **58,** 438-458.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston, MA: Authors Cooperative, Inc.

Sidman, M. (1997a). Equivalence: A theoretical or a descriptive model? *Mexican Journal of Behavior Analysis,* **23,** 125-145.

Sidman, M. (1997b). Equivalence relations [commentary]. *Journal of the Experimental Analysis of Behavior,* **68,** 258-266.

Sidman, M., & Cresson, O. (1973). Reading and the cross-modal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency,* **77,** 515-523.

Sidman, M., Cresson, O., & Willson-Morris, M. (1974). Acquisition of matching-to-sample via mediated-transfer. *Journal of the Experimental Analysis of Behavior,* **22,** 261-273.

Sidman, M., Kirk, B., & 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., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & 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.

Sidman, M., & Tailby, W. (1982). Conditional discrimination *vs.* matching-to-sample: An extension of the testing paradigm. *Journal of the Experimental Analysis of Behavior,* **37,** 5-22.

Sidman, M., Willson-Morris, M., & 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., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior,* **52,** 261-274.

Sigurdardottir, Z. G. (1992). *Establishing classes of Icelandic nouns using a stimulus equivalence paradigm.* Unpublished doctoral dissertation, Northeastern University, Boston.

Sigurdardottir, Z. G., Green, G., & Saunders, R. R. (1990). Equivalence classes generated by sequence training. *Journal of the Experimental Analysis of Behavior,* **53,** 47-63.

Skinner, B. F. (1957). *Verbal Behavior.* New York: Appleton-Century-Crofts.

Skinner, B. F. (1969). *Contingencies of Reinforcement: A Theoretical Analysis.* New York: Appleton-Century-Crofts.

Smagorinsky, P. (1989). The reliability and validity of protocol analysis. *Written Communication,* **6,** 463-479.

Smagorinsky, P. (1998). Thinking and speech and protocol analysis. *Mind, Culture, and Activity,* **5,** 157-177.

Smeets, P. M., & Barnes, D. (1995). Emergent simple discrimination via transfer from differentially reinforced S+ stimuli: A further test of the stimulus-response interaction model. *The Quarterly Journal of Experimental Psychology,* **48B,** 329-345.

Smeets, P. M., & Barnes, D. (1997). Emergent conditional discriminations in children and adults: Stimulus equivalence derived from simple discriminations. *Journal of Experimental Child Psychology,* **66,** 64-84.

Smeets, P. M., Barnes, D., & Luciano, C. M. (1995). Reversal of emergent simple discrimination in children: A component analysis. *Journal of Experimental Child Psychology,* **60,** 327-343.

Smeets, P. M., Barnes, D., & Roche, B. (1997). Functional equivalence in children: Derived stimulus-response and stimulus-stimulus relations. *Journal of Experimental Child Psychology,* **66,** 1-17.

Smeets, P. M., Leader, G., & Barnes, D. (1997). Establishing stimulus classes in adults and children using a respondent-type training procedure: a follow-up study. *The Psychological Record,* **47,** 285-308.

Smeets, P. M., Schenk, J. J., & Barnes, D. (1994). Establishing transfer from identity to arbitrary matching tasks via complex stimuli under testing conditions: A follow-up study. *The Psychological Record,* **44,** 521-536.

Smeets, P. M., & Striefel, S. (1994). Matching to complex stimuli under non-reinforced conditions: Errorless transfer from identity to arbitrary matching tasks. *The Quarterly Journal of Experimental Psychology,* **47B,** 39-62.

Smith, A. B., Dickins, D. W., & Bentall, R. P. (1996). The role of interfering stimulus names in the emergence of equivalence relations II: The effects of interfering tasks prior to and after tests for equivalence relations. *The Psychological Record,* **46,** 109-130.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Learning and Memory,* **6,** 174-215.

Solomon, P. (1995). The think-aloud method: A practical guide to modelling cognitive processes. *Information Processing and Management,* **31,** 906-907.

Spencer, T. J., & Chase, P. N. (1996). Speed analyses of stimulus equivalence. *Journal of the Experimental Analysis of Behavior,* **65,** 643-659.

Spradlin, J. E., Cotter, V. W., & 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.

Spradlin, J. E., & Dixon, M. H. (1976). Establishing conditional discriminations without direct training: Stimulus classes and labels. *American Journal of Mental Deficiency,* **80,** 555-561.

Spradlin, J. E., & Saunders, R. R. (1984). Behaving appropriately in new situations: A stimulus class analysis. *American Journal of Mental Deficiency,* **88,** 574-579.

Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification versus comparison classification. *Analysis and Intervention in Developmental Disabilities,* **6,** 41-58.

Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior,* **56,** 519-555.

Stemmer, N. (1996). Listener behavior and ostensive learning [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 247-249.

Stikeleather, G., & Sidman, M. (1990). An instance of spurious equivalence relations. *The Analysis of Verbal Behavior,* **8,** 1-11.

Stoddard, L. T., & McIlvane, W. J. (1986). Stimulus control research and developmentally disabled individuals. *Analysis and Intervention in Developmental Disabilities,* **6,** 155-178.

Stromer, R. (1986). Control by exclusion in arbitrary matching-to-sample. *Analysis and Intervention in Developmental Disabilities,* **6,** 59-72.

Stromer, R. (1996). On the experimental analysis of naming and the formation of stimulus classes [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 250-252.

Stromer, R., & Mackay, H. A. (1990). Arbitrary stimulus relations and delayed identity matching to complex samples. *Experimental Analysis of Human Behavior Bulletin,* **8,** 22-25.

Stromer, R., & Mackay, H. A. (1992a). Delayed constructed-response identity matching improves the spelling performance of students with mental retardation. *Journal of Behavioral Education,* **2,** 139-156.

Stromer, R., & Mackay, H. A. (1992b). Spelling and emergent picture-printed word relations established with delayed identity matching to complex samples. *Journal of Applied Behavior Analysis,* **25,** 893-904.

Stromer, R., & Mackay, H. A. (1993). Delayed identity matching to complex samples: Teaching students with mental retardation spelling and the prerequisites for equivalence classes. *Research in Developmental Disabilities,* **14,** 19-38.

Stromer, R., & Mackay, H. A. (1996). Naming and the formation of stimulus classes. In T. R. Zentall and P. M. Smeets (Eds.), *Advances in Psychology:*

117. Stimulus class formation in humans and animals (participant. 221-252). Amsterdam: Elsevier.

Stromer, R., Mackay, H. A., Howell, S. R., McVay, A. A., & Flusser, D. (1996). Teaching computer-based spelling to individuals with developmental and hearing disabilities: Transfer of stimulus control to writing tasks. *Journal of Applied Behavior Analysis,* **29,** 25-42.

Stromer, R., Mackay, H. A., McVay, A. A., & Fowler, T. (1998). Written lists as mediating stimuli in the match-to-sample performances of individuals with mental retardation. *Journal of Applied Behavior Analysis,* **31,** 1-19.

Stromer, R., Mackay, H. A., & Remington, B. (1996). Naming, the formation of stimulus classes, and applied behavior analysis. *Journal of Applied Behavior Analysis,* **29,** 409-431.

Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education,* **3,** 225-256.

Stromer, R., McIlvane, W. J., Dube, W. V., & Mackay, H. A. (1993). Assessing control by elements of complex stimuli in delayed matching to sample. *Journal of the Experimental Analysis of Behavior,* **59,** 83-102.

Stromer, R., McIlvane, W. J., & Serna, R. W. (1993). Complex stimulus control and equivalence. *The Psychological Record,* **43,** 585-598..

Stromer, R., & 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.

Stromer, R., & Stromer, J. B. (1989). Childrens' identity matching and oddity: Assessing control by specific and general sample-comparison relations. *Journal of the Experimental Analysis of Behavior,* **51,** 47-64.

Stromer, R., & Stromer, J. B. (1990a). Matching to complex samples: Further study of arbitrary stimulus classes. *The Psychological Record,* **40,** 505-516.

Stromer, R., & Stromer, J. B. (1990b). The formation of arbitrary stimulus classes in matching to complex samples. *The Psychological Record,* **40,** 51-66.

Svenson, O. (1985). Protocol analysis: Verbal reports as data. *British Journal of Psychology,* **76,** 277-278.

Tena, R. O., & Velazquez, H. A. (1997). An exploratory study on teaching the reading of musical notes through the use of conditional discrimination. *Revista Mexicana de Psicologia,* **14,** 13-29.

Tierney, K. J., De Largy, P., & Bracken, M. (1995). Formation of an equivalence class incorporating haptic stimuli. *The Psychological Record,* **45,** 431-437.

Tomonaga, M. (1993). Tests for control by exclusion and negative stimulus relations of arbitrary matching to sample in a "symmetry-emergent" chimpanzee. *Journal of the Experimental Analysis of Behavior,* **59,** 215-229.

Tomonaga, M. (1999). Establishing functional classes in a chimpanzee (*pan troglodytes*) with a two-item sequential-responding procedure. *Journal of the Experimental Analysis of Behavior,* **72,** 57-79.

Touchette, P. E. (1971). Transfer of stimulus control: Measuring the moment of transfer. *Journal of the Experimental Analysis of Behavior,* **15,** 347-354.

Trigo, E., Martinez, R., & Moreno, R. (1995). Rule performance and generalization in a match-to-sample task. *The Psychological Record,* **45,** 223-240.

Urcuioli, P. J., & Zentall, T. R. (1993). A test of comparison-stimulus substitutability following one-to-many matching by pigeons. *The Psychological Record,* **43,** 745-759.

Urcuioli, P. J., Zentall, T. R., Jackson-Smith, P., & Steirn, J. N. (1989). Evidence for common coding in many-to-one matching: Retention, intertrial interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes,* **15,** 264-273.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes,* **14,** 36-42.

Vaughan, W., Jr. (1989). Reply to Hayes. Journal of the Experimental Analysis of Behavior, **51,** 397.

Venkatesan, M. (1986). Protocol analysis: Verbal reports as data. *Journal of Marketing Research,* **23,** 306-307.

Vygotsky, L. S. (1962). *Thought and Language.* Cambridge, MA: MIT Press.

Whitehurst, G. J. (1996). On the origins of misguided theories of naming and other symbolic behavior [commentary]. *Journal of the Experimental Analysis of Behavior,* **65,** 255-259.

Wilson, K. G., & Hayes, S. C. (1996). Resurgence of derived stimulus relations. *Journal of the Experimental Analysis of Behavior,* **66,** 267-281.

Wilson, T. D. (1994). The proper protocol: Validity and completeness of verbal reports. *Psychological Science,* **5,** 249-252.

Wulfert, E., Dougher, M. J., & Greenway. D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. *Journal of the Experimental Analysis of Behavior,* **56,** 489-504.

Wulfert, E., Greenway, D. E., & Dougher, M. J. (1994). Third-order equivalence classes. *The Psychological Record,* **44,** 411-439.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior,* **50,** 125-144.

Yamamoto, J., & Asano, T. (1995). Stimulus equivalence in a chimpanzee (*Pan troglodytes*). *The Psychological Record,* **45,** 3-21.

Zentall, T. R. (1998). Symbolic representation in animals: Emergent stimulus relations in conditional discrimination learning. *Animal Learning and Behavior,* **26,** 363-377.

Zentall, T. R., & Urcuioli, P. J. (1993). Emergent relations in the formation of stimulus classes by pigeons. *The Psychological Record,* **43,** 795-810.

Zygmont, D. M., Lazar, R. M., Dube, W. V., & McIlvane, W. J. (1992). Teaching arbitrary matching via sample stimulus-control shaping to young children and mentally retarded individuals: A methodological note. *Journal of the Experimental Analysis of Behavior,* **57,** 109-117.