Title	Testing and developing procedures for assessing and training hierarchical classification skills in young children using relational frame theory
Author(s)	Mulhern, Teresa
Publication Date	2018-02-22
Item record	http://hdl.handle.net/10379/7159

Testing and Developing Procedures for Assessing and Training Hierarchical Classification Skills in Young Children Using Relational Frame Theory

Teresa Mulhern

BA (Hons) Psychology

MSc Applied Behaviour Analysis

Dissertation submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Applied Behaviour Analysis

Supervisor: Dr Ian Stewart

School of Psychology

National University of Ireland, Galway

October, 2017

This research was supported by the National University of Ireland, Galway Doctoral Research Scholarship Scheme.

Table of Contents

Declaration Regarding the Work

I, the **Candidate,** certify that the Thesis is all my own work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

This thesis is the result of my own investigations, except where otherwise stated. The following chapter includes work conducted in collaboration with other students:

- Chapter 6 includes work conducted in collaboration with Siri Ming, Ph.D – Study 7 (Thesis titled *Assessing and Training Early Emergent Derived Relational Responding in Children with Autism* submitted September 2015).
    - Co-designed the intervention package for Studies 6 and 7, collected data and provided class inclusion training for Study 6 and collected IOA data for Study 7.
- Chapter 6 includes work conducted in collaboration with Patrycja Zagrabska, ABA MSc Student (Study 8).
    - Co-designed the intervention package for Study 8 and collected IOA data.

Signed: ________________________ Date: ____________________

Publications and Conference Presentations Resulting from this Thesis

Publications

Mulhern, T., Stewart, I., & McElwee, J. (2017). Investigating relational framing of categorization in young children. *The Psychological Record*. DOI: 10.1007/s40732-017-0255-y

Ming, S., Mulhern, T., Stewart, I., Moran, L., & Bynum, K. (2017). Training class inclusion responding in typically-developing children and individuals with autism. *Journal of Applied Behavior Analysis*. DOI: 10.1002/jaba.429

Manuscripts in Preparation/Under Submission

Mulhern, T., Stewart, I., & McElwee, J. (under submission). Facilitating relational framing of classification in young children.

Zagrabska, P., Mulhern, T., & Stewart, I. (in preparation). Testing and training class inclusion responding based on a relational frame theory approach with individuals with autism.

Conference Presentations

Mulhern, T. & Stewart, I. (2014, September). Training class inclusion responding in young children. Paper presented at the annual conference of the European Association for Behaviour Analysis, Stockholm, Sweden.

Mulhern, T. & Stewart, I. (2014, December). Teaching class inclusion responding to young children. Paper presented at the Acceptance and Commitment Therapy, Contextual and Behavioral Science conference, Dublin.

Ming, S., Mulhern, T., Moran, L., & Stewart, I. (2015, March). Training class inclusion skills with typically developing children and children with autism. Paper presented at the Association of Professional Behaviour Analysis conference, Seattle.

Mulhern, T., Ming, S., Moran, L., & Stewart, I. (2015, April). Training class inclusion responding in young children and individuals with autism. Paper presented at the Division of Behaviour Analysis conference, Galway.

Mulhern, T., & Stewart, I. (2016, September). Investigating containment and hierarchical relational responding in young children. Paper presented at the European Association of Behaviour Analysis, Sicily.

Mulhern, T., Ming, S., Moran, L., & Stewart, I. (2016, September). Teaching class inclusion to typically developing children and individuals with autism. Paper presented at the European Association of Behaviour Analysis, Sicily.

Mulhern, T., & Stewart, I. (2016, November). Investigating containment and hierarchical relational responding in young children. Paper presented at the Acceptance and Commitment Therapy, Contextual and Behavioural Science Conference, Edinburgh.

Mulhern, T., & Stewart, I. (2017, April). Facilitating containment and hierarchical relational responding repertoires in young children. Paper presented at the Experimental Analysis of Behaviour Group conference, London.

Mulhern, T., Zagrabska, P., Ming, S., & Stewart, I. (2017, May). Training class inclusion in individuals with autism. Paper presented at the Psychological Society of Ireland, Division of Behaviour Analysis conference, Trinity College Dublin.

Mulhern, T., & Stewart, I. (2017, May). Assessing and training categorization repertoires in young children. Paper presented at the Psychological Society of Ireland, Division of Behaviour Analysis conference, Trinity College Dublin.

Abstract

The current thesis aimed to conceptualise, assess and train features of hierarchical classification from a behaviour analytic, or more specifically a Relational Frame Theory (RFT), perspective. Classification refers to grouping stimuli according to shared physical or functional characteristics (Barnes-Holmes, Dymond, & O'Hora, 2001). Hierarchical classification is a more complex form of classification, whereby classes themselves are classified as members of other classes. For example, a "budgie" is a member of the class of "bird", and the class of "bird" is a member of the class of "animal". Classification can be conceptualised as involving particular types of framing. RFT sees containment (A is in B; B contains A) and hierarchical (A is a type of B; B is a class containing A) relational responding as core repertoires for categorisation. As such, RFT regards both containment and hierarchical relational responding as core repertoires for categorisation.

Study 1 employed a protocol to assess these repertoires in young typically developing children (n = 50; 3 – 8 years). The relational protocol developed for this study assessed mutual entailment, combinatorial entailment, and transformation of function in three relational domains including non-arbitrary and arbitrary containment relations and arbitrary hierarchical relations. The study also aimed to correlate relational framing performance with linguistic and cognitive potential as measured by standardised instruments including the Peabody Picture Vocabulary Test 4th edition (PPVT-4), Stanford Binet 5th Edition (SB5) and Children's Category Test (CCT) in young children. Results provided data concerning the acquisition of relational categorisation skills across childhood. The research also showed strong positive correlations between relational performance and linguistic (PPVT-4), categorisation (CCT), and cognitive performance (SB5).

Study 2 trained arbitrary containment relational responding in a 5-year-old using a multiple baseline across components design at varying levels of complexity (i.e., mutual entailment, combinatorial entailment, and transformation of function). Prior to her inclusion within the relational training study, the participant was first screened for suitability using a non-arbitrary containment relational repertoire assessment. The results indicated that she had a sufficiently strong non-arbitrary containment repertoire to be included within the research. Training involved reinforcement, contingent feedback based on arbitrary containment contextual cues and correction procedures. Correct responding increased from baseline to criterion performance in each component skill upon intervention, and generalisation and maintenance were also observed. Study 3 extended this research by training arbitrary containment relational responding in three 5-year-olds using a combined multiple baseline (across both components and participants) design. Results demonstrated the effectiveness of training for increasing responding in accordance with arbitrary containment and generalisation was also observed. Participants were also assessed pre- and post-training for their receptive language scores (using the PPVT-4), categorisation repertoire (using the CCT), and class inclusion responding to determine the impact of training on these untrained variables. The participants' scores were compared to three 5-year-olds who were not exposed to training at both time points (i.e., pre- and post-training). The results indicated that arbitrary containment training resulted in increases in class inclusion responding, receptive language and categorisation scores.

Study 4 trained arbitrary hierarchical relational responding in a 6-year-old using a multiple baseline across components design at varying levels of complexity (i.e., mutual entailment, combinatorial entailment, and transformation of function). Training involved reinforcement, contingent feedback based on arbitrary hierarchical contextual cues and correction procedures. Correct responding increased from baseline to criterion performance in each component skill upon intervention, and generalisation and maintenance were also observed. Study 5 extended this research by training arbitrary hierarchical relational responding in three 6-year-olds using a combined multiple baseline (across both components and participants) design. Results demonstrated the effectiveness of training for increasing responding in accordance with arbitrary containment and generalisation was also observed. Participants were also assessed pre- and post-training for their receptive language scores (using the PPVT-4), categorisation repertoire (using the CCT), and class inclusion responding to determine the impact of training on these untrained variables. The participants' scores were compared to three 6-year-olds who were not exposed to training at both time points (i.e., pre- and post-training). The results indicated that arbitrary containment training resulted in increases in class inclusion responding, receptive language and categorisation scores.

Studies 6 - 8 used procedures based on RFT to assess and train Piagetian class inclusion type responding in both typically developing and developmentally delayed children. In Study 6, typically developing 3-year-olds were exposed to multiple exemplar non-arbitrary class inclusion training in the presence of contextual cues in the context of a multiple baseline across participants design. Correct responses were reinforced using a token economy. Class inclusion responding increased significantly relative to baseline contingent on the introduction of the intervention. Generalisation and maintenance of class inclusion responding was also observed. Study 7 extended the research by applying a slightly modified version of the training protocol to children with a diagnosis of autism. Intervention was successful in this case also though on a somewhat more limited basis. Finally, Study 8 employed a modified version of the training protocol to individuals with autism and a secondary diagnosis within the context of a concurrent multiple baseline design. The results indicated an increase in class inclusion responding for all participants relative to baseline, in addition to generalisation of class inclusion responding to untrained categories and maintenance of treatment gains. Findings and recommendations for future work are discussed.

List of Tables

List of Figures

List of Appendices

Abbreviations

AARR	Arbitrarily Applicable Relational Responding
ABA	Applied Behaviour Analysis
ASD	Autism Spectrum Disorder
CCT	Children's Category Test
CE	Combinatorial Entailment
CI	Class Inclusion
EVT-2	Expressive Vocabulary Test, 2nd edition
FR	Fixed Ratio
FSIQ	Full Scale Intelligence Quotient
IOA	Inter-Observer Agreement
ME	Mutual Entailment
MET	Multiple Exemplar Training
MTS	Match-To-Sample
NAARR	Non-Arbitrary Applicable Relational Responding
NMT	Network Model Theory
NT	Neuro-typical
NVIQ	Non-Verbal Intelligence Quotient
PECS	Picture Exchange Communication System
PPVT-4	Peabody Picture Vocabulary Test, 4th edition
PPVT-4 AE	Peabody Picture Vocabulary Test, 4th edition – Age Equivalence
REP	Relational Evaluation Procedure
ToF	Transformation of Function

ToF-C	Transformation of Function of Combinatorially Entailed Relations
ToF-M	Transformation of Function of Mutually Entailed Relations
VIQ	Verbal Intelligence Quotient
WAIS-III	Wechsler Adult Intelligent Scale, 3rd edition
WISC-III	Wechsler Intelligence Scale for Children, 3rd edition
WPPSI-R	Wechsler Preschool and Primary Scale of Intelligence-Revised
RFT	Relational Frame Theory
SB5	Stanford Binet, 5th edition

Glossary of Terms

Arbitrarily Applicable Relational Responding	Learned relational responding under the control of arbitrary contextual cues (e.g., "type of", "contains", etc.) the relata involved. This form of responding is not reliant on the physically properties of the relata nor on direct experience with them.
Non-Arbitrary Responding	Responding is based on the physical properties of stimuli (e.g., colour, height, etc.)
Mutual Entailment	A defining feature of relational framing that refers to its fundamental bi-directionality. Mutual entailment applies when in a given context A is related in a characteristic way to B, and as a result B is now related in another characteristic way to A.
Combinatorial Entailment	A defining feature of relational framing that refers to the derivation of a new relation based on the combination of previously acquired relations. For example, combinatorial entailment applies when in a given context A is related in a characteristic way to B, and to C, and as a result a relation between B and C is mutually entailed.
Transformation of Stimulus Function	A defining feature of relational framing that refers to the modification of the stimulus functions of relata based on their participation in relational frames.
Derived Relational Responding	The derivation of new untaught relations from previously acquired (i.e., either directly taught or derived) relations.

Chapter 1

Hierarchical Classification:

A Review of Mainstream and Behaviour Analytic Research

The primary focus of the behaviour analytic and more specifically Relational Frame Theory-based research reported in this thesis is the assessment and training of relational repertoires involved in hierarchical classification responding in young children including both typically developing children as well as individuals with a diagnosis of autism spectrum disorder (ASD).

Classification has been defined as grouping stimuli into distinct classes or categories based on common functions or physical features, while a class or category is a collection of stimuli that cohere in this manner (Astley, Peissig, & Wasserman, 2001; Zentall, Galizio, & Critchfield, 2002). In hierarchical classification, which is a relatively advanced form of classification, classes are themselves categorised into higher order classes (Greene, 1994). An example of this would be classifying “budgie” into the category “bird”, while classifying “bird” into the category “animal”.

Classification is a fundamentally important repertoire with regard to understanding and navigating the human environment (Markman, 1989). Categorising stimuli at various levels of complexity allows an individual to systematically adjust the functions of those stimuli so as to guide their future behaviour effectively (Bornstein & Mash, 2010; Furrer & Younger, 2008; Gelman, 1988; Kalish & Gelman, 1992; Lin & Murphy, 2001; Proffitt, Coley & Medin, 2000). Hierarchical categorisation is particularly important in this regard and not just at a personal but at a societal level. Hierarchical categorisation is critical for scientific thinking for example. For instance, the production of taxonomies (e.g., within biology or chemistry) is heavily dependent on hierarchical organisation. As such, teaching children to respond scientifically requires teaching them to respond in accordance with the logic of hierarchical classification.

Thus, hierarchical classification is a fundamentally important repertoire for all children to be taught effectively and thus it behoves us as behavioural scientists to develop an understanding of the key processes involved so as to facilitate effective training of this repertoire. That is a core focus of this thesis.

The current (introductory) chapter will review research on classification and especially hierarchical classification from a number of different perspectives. It will start by presenting theoretical underpinnings and empirical work from researchers within mainstream cognitive and cognitive developmental arenas. It will then proceed to examine behaviour analytic research on classification and hierarchical classification that occurred prior to Relational Frame Theory-based work, which is covered in Chapter 2.

Cognitive research

The majority of empirical research into classification including hierarchical classification has originated from the domain of cognitive psychology (e.g., Blewitt, 1994; Deneault & Ricard, 2006; Gelman, 1988; Greene, 1994; Inhelder & Piaget, 1964). In what follows I will discuss a prominent cognitive approach to hierarchical classification specifically.

Philosophical Assumptions. Before starting, I will first briefly discuss philosophical differences between the cognitive approach and the behaviour analytic approach. Cognitive researchers typically assume that basic and complex human behaviour is mediated by mental mechanisms and that the key goal of psychology is the identification of those mechanisms and the conditions under which they operate. In contrast, behaviour analysts typically explain human behaviour as the causal product of (past and present) environmental variables and assume that the key goal of psychology is the identification of these variables so as to allow both prediction and influence over human behaviour. As such, with regard to classification just as with other behaviour patterns, cognitive researchers hypothesise and test mechanisms (typically based on the metaphor of the mind as computer or information processor) thought to mediate this behaviour while the behaviour analytic- (and more specifically RFT-) based work in this thesis is concerned to highlight the environmental variables required for successfully influencing classification behaviour and more specifically required for establishing this behaviour in various populations (and for this reason, an important phase of the work reported in this thesis involved training various groups in classification-relevant repertoires). Despite these philosophical differences, it makes sense that approaches focused on the same behavioural phenomenon can learn from each other. Nevertheless, the reader should be aware that findings are always interpreted in light of the awareness of this difference in implicit assumptions.

With regard to categorisation specifically, cognitive researchers have sought to discover how the mind works to group stimuli with either physical or functional commonalities and to encode these as mental representations or concepts. Within this paradigm, it is argued that the ability to classify acts as a means to simplify the environment, reduce memory burden, and help to both save and retrieve information efficiently and effectively. From this perspective the categorisation of stimuli provides us with important information regarding these stimuli, including features or characteristics that may be held by them by virtue of their category placement (Baldwin, Markman, & Melartin, 1993; Greco, Hayne,

& Rovee-Collier, 1990; Mandler, 2000; Needham, Cantlon, & Ormsbee Holley, 2006). As such, categorisation is seen as an essential process in which meaning is assigned to novel stimuli; providing us with the ability to then select appropriate behavioural responses to these stimuli based upon their category placement (Freedman & Assad, 2006). The variability and flexibility that individuals display in categorisation with respect to their larger knowledge system (e.g., in developing analogies, forming inferences, or constructing theories) is seen as demonstrating that classification is a process that is innately linked to virtually all aspects of cognition (Gelman & Meyer, 2011).

Network Model Theory. The focus of the current thesis is hierarchical classification specifically and hence I will discuss a relatively recent and prominent cognitive account of hierarchical classification referred to as Network Model Theory (NMT; e.g., Collins & Quillian, 1969; Gluck & Bower, 1988; Love, Medin & Gureckis, 2004). According to NMT, the mental system involves a network of hierarchically organised nodes representing classes/concepts. Nodes perform one of two roles: ‘type’ nodes save information relevant to a class (e.g., ‘Cats have fur’), while the alternative nodes save facts for an example of that class (e.g., ‘I have a black cat’; Murphy, 2002). The information saved at each node is not duplicated within this network, which is consistent with the principle of cognitive economy. For example, the distinguishing characteristics of a category (e.g., reptiles are cold-blooded) are saved at the node that represents that specific category, and are not saved at an alternative node (e.g., at the node for lizards). In addition to this, each node has a superordinate node that determines class membership (Best, 1995), and a subordinate node within the hierarchical class network. Superordinate nodes are classes (e.g., reptile) that incorporate other classes or subordinate nodes (e.g., lizards).

Properties of Hierarchical Networks. Relations between nodes in the hierarchical network are in accordance with the principles of set inclusion (Murphy, 2002). For example, the set of animals includes the set of reptiles, which includes the set of lizards. Set-inclusion in turn is defined by three properties; transitive class containment, asymmetric class containment, and unilateral property induction (Barrouillet, 1996; Billows, 1975; Blewitt, 1994; Bruner, Goodnow & Austin, 1956; Collins & Quillian, 1969; Deneault & Ricard, 2006; Diesendruck & Shatz, 2001; Gelman, 1988; Greene, 1994; Inhelder & Piaget, 1964; Murphy, 2002; Winer, 1980).

Transitive class containment is the idea that if X (subordinate) is a member of class Y (intermediate) and Y is a member of class Z (superordinate), then X (subordinate) is a member of Z (superordinate). For example, all lizards (subordinate) are reptiles (intermediate), and all reptiles (intermediate) are animals (superordinate); as such, all lizards (subordinate) are animals (superordinate). Asymmetrical class containment is the idea that if class X contains class Y, then class Y cannot contain class X. For example, the class of reptiles includes the class of snakes, and so the class of snakes cannot include the class of reptiles. The third defining feature of hierarchical classes is unilateral property induction, which refers to the asymmetrical transfer of properties from higher (e.g., superordinate) categories to lower (e.g., subordinate) categories but not the other way round. Thus, the properties of a superordinate category (e.g., reptile) are explicitly shared with all members of subordinate classes (e.g., lizards and snakes), but the properties of the subordinate classes are not necessarily shared with members of the superordinate class (Johnson, Scott, & Mervis, 1997; Murphy, 2002). For example, a feature of the class of “reptile” is that they are cold-blooded; therefore, all members and subclasses of “reptile” have this particular characteristic. However, while the class “lizard”, which is a subclass of “reptile”, has the property of “four legs”, not all “reptiles” have four legs.

NMT postulates that when an individual encounters a novel stimulus it leads to the activation of nodes at different levels within the hierarchical network and information is transported along these pathways consistent with the previously outlined features that characterise the set inclusion relation, allowing an individual to classify a stimulus (Collins & Quillian, 1969; Murphy, 2002). Results from a number of studies have supported NMT. This includes work on response latencies as a function of nodal distance between hierarchically related semantic categories. To explain, if stimuli are in a hierarchically structured network, as NMT suggests, then this might point to predictable differences in nodal distance and hence to predictable differences in response latency in priming paradigms. In support of this, research employing sentence verification tasks, for instance, has found that individuals are faster to identify a stimulus (e.g., ‘cobra’) as a member of an intermediate class (e.g., ‘snake’) (which is one node away) than as a member of a superordinate class, such as “reptile” (which is two nodes away) (Collins & Quillian, 1969, 1972; Murphy, 2002; Rips, 1975; Rips, Shoben & Smith, 1973).

Apart from work testing NMT as a description of the mental system underlying hierarchical categorisation, there is also research examining people’s responding in accordance with hierarchical semantic organisation more generally. For example, categories contained in memory can vary in terms of their complexity and can include numerous category levels, depending on a person's history and expertise in relation to specific categorical taxonomies (Coley, Medin, & Atran, 1997; López, Atran, Coley, Medin & Smith, 1997; Medin, Lynch, Coley & Atran, 1997; Murphy, 2002; Tanaka & Taylor, 1991). However, Eysenck and Keane (2000), for example, have indicated that humans typically categorise to three levels of generality, namely the superordinate, intermediate (basic), and subordinate levels (Mervis & Rosch, 1981; Rosch, 1978). Further to this, people are more likely to classify stimuli at the basic level than at either the superordinate or subordinate levels (Berlin, 1972, 1992; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Ross & Murphy, 1999). This is suggested to be because such classes mirror natural discontinuities within the environment and also have maximum 'cue validity', which is the quantity of shared and defining features of a specific class and the optimal level of differentiation from other classes.

Summary and Conclusion. The current broad consensus from cognitive psychologists based on the most recent empirical research, is that categories are stored hierarchically in memory and are characterised by features of set inclusion previously described, specifically, transitive class containment, asymmetrical class containment, and unilateral property induction (Murphy, 2002). As will be seen, this perspective has been influential in shaping some aspects of the RFT-based approach which is the one adopted within the current research programme. In the current section, I have reviewed one relatively recent empirically based cognitive theory which is representative of cognitive theory on hierarchical classification more generally. In the next section, I turn to examination of cognitive developmental research.

Cognitive Developmental Research

Cognitive developmental psychologists are concerned with describing changes in cognition over the lifespan but in particular within childhood and adolescence.

Philosophical Assumptions. This area has been influenced by both organicist-developmentalist and mechanistic-cognitivist assumptions. As described previously, mechanistic-cognitivism assumes that basic and complex human behaviour is mediated by mental mechanisms and that the key goal of psychology is the identification of those mechanisms. This is the predominant paradigm within psychological research and up to this point in the current chapter we have reviewed research from this perspective on hierarchical classification in adults. This paradigm is influential also in developmental research, and in particular with respect to the development of cognitive skills such as hierarchical classification. At the same time, developmental research is also influenced by organicist-developmentalist assumptions which are grounded in aspects of organismic growth including, for example, maturation over an invariant sequence of developmental stages. Piaget's account of cognitive development provides the archetypal example of this perspective. For example, in the Piagetian account, cognition naturally develops in an ordered and invariant sequence of stages, with each stage being identifiable by certain prototypical features.

The behaviour analytic approach adopted within this thesis rejects the concept of invariant stages just as it rejects the mentalism of the cognitive approach. In addition, both mechanistic-cognitivism and organicist-developmentalism feature a correspondence-based truth criterion which contrasts with the pragmatic truth criterion of behaviour analysis which seeks prediction and influence over behaviour. Hence the behaviour analytic approach adopted herein has different basic assumptions with regard to development than those adopted in much of the previous work conducted within the developmental realm. Nevertheless, once again, reviewing the work of previous researchers can yield important guidance for the behaviour analytic work conducted as part of the current programme of research.

Research on Categorisation. Cognitive developmental research has provided a considerable amount of data on the development of categorisation in early infancy and childhood (e.g., Blewitt, 1994; Bornstein, Arterberry, & Mash, 2010; Deneault & Ricard, 2006; Gelman, 1988; Greene, 1994; Inhelder & Piaget, 1964; Johnson, et al., 1997). This includes work on both the development of more basic skills (e.g., investigating the ages at which, for example, children become capable of distinguishing between various groups of objects) as well as more advanced ones (e.g., assessing when children start to show responding consistent with properties of hierarchical classification previously discussed such as transitive class containment, asymmetric class containment, and unilateral property induction).

The acquisition of categorisation skills including hierarchical categorisation is a gradual process spanning childhood and adolescence. The cognitive developmental approach, which merges both cognitive and developmental perspectives on classification, contends that such changes to an individual's repertoire occur as a result of qualitative changes in mental representations (Quinn & Eimas, 1997), which occur as a result of the shift from perceptual to conceptual classification (Murphy, 2002; Quinn & Eimas, 1997). Perceptual classification refers to the grouping of stimuli based on their formal or physical features (Fields & Reeve, 2000). Children start by learning this more basic type of classification based on their experience of actual physical items and their ability to compare and contrast those items. This then provides the basis for the development of more complex conceptual classification skills in which an individual can make classification responses based on inferential knowledge and between-category reasoning (Deneault & Ricard, 2006; Gelman, 2003; Mandler, 2000; Murphy, 2002; Quinn & Eimas, 1997). At the core of this is ‘hierarchical’ classification (Blewitt, 1994; Deneault & Ricard, 2006; Diesendruck & Shatz, 2001; Inhelder & Piaget, 1964; Vygotsky, 1962).

Blewitt: Theory & Research. Blewitt has conducted a considerable amount of research on the development of conceptual and in particular hierarchical classification. Blewitt (1989, 1994) suggests that there are two skills needed for comprehension of conceptual hierarchies including the ability to (i) develop categories at varying levels of generality, and (ii) incorporate the same objects into multiple categories. She reports that while two-year-olds seem to have the first skill, the second is acquired later in childhood. Other researchers have indicated that the ability to form accurate inferences with respect to hierarchically structured categories is not acquired until the age of five (Inhelder & Piaget, 1964; Winer, 1980).

Blewitt (1989, 1994) has suggested four levels in the development of classification ability in children’s preschool years. In Level 1 (< 2 years), children show the emergence of a preliminary categorisation skill, namely, the ability to form categories at varying levels of generality. For example, in this stage, a child can label a stimulus as both a snake and an animal, but might not agree that a snake is an animal. Once children reach Level 2 (2-5 years), they can enter the same objects into multiple categories at varying levels of generality (Blewitt, 1989; Blewitt, 1994). For example, a child will now consider a snake to be a member of the categories “animal”, “pet” and “reptile” simultaneously, which Blewitt sees as suggesting that they are now capable of including the class of snakes in the category of animals (Blewitt, 1994). Blewitt (1989) argues that the presence of Level 2 behaviour suggests that a child is now establishing a conceptual framework and is also starting to mentally represent hierarchically related classes.

The knowledge and skills acquired in Levels 1 and 2 are seen as prerequisites for the more complex repertoires that emerge in Levels 3 and 4 (Blewitt, 1994). At Level 3 (5-7 years) children begin to form inferences from hierarchies and to reason about classes without referring to the surface characteristics or formal features of the stimuli in question. They begin to acquire explicit inter-category knowledge and start to show transitive class containment. For example, they can reason, “If snakes are animals, and cobras are snakes, then cobras are animals”. In the final stage, Level 4 (7-11 years), they become capable of asymmetrical inclusion and unilateral induction relations; for example, they can reason that “All animals breathe air. Snakes are animals, so snakes breathe air. However, while all snakes have scales, not all animals have scales.” As a result of progressing to Level 4, children become capable of performing quantitative inferences about relative category sizes within a categorical hierarchy and as such demonstrate accurate performances on tasks like Inhelder and Piaget’s (1964) class inclusion task.

Class Inclusion. The Piagetian class inclusion task (e.g., Piaget, 1952) is a quite well-known cognitive developmental task. It was used by Piaget and others to determine if a child had reached the “concrete operational” stage of Piaget’s stage-based account of cognitive development, which was theorised to occur by age seven or eight (Brainerd, 1974). From the current perspective, it can be seen as a relatively important marker of hierarchical classification skill. In the class inclusion task, a child is first shown an array of stimuli in a particular class that includes two different subclasses with a greater quantity of one than the other. They are then asked whether there are more members of the more populous of the two subclasses or more members of the entire class. For example, they might be shown dogs and cats with more dogs than cats present and asked “Are there more dogs or are there more animals?” A key requirement for correct responding is to be able to respond to particular stimuli as simultaneously belonging both to a class as well as to a subclass contained within the larger class.

Cognitive developmental psychologists argue that in order to perform class inclusion correctly, a number of mental skills are required. First, the child must “conserve” the superordinate class of animals and the subordinate classes of dogs and cats in working memory (Inhelder & Piaget, 1964; Thomas & Horton, 1997). They must then add the subclasses (e.g., dogs + cats = animals), which is considered evidence for transitive class containment, as in order to do so they must understand that individual stimuli are contained within their respective class (e.g., each of the dogs is in the ‘dogs’ class), that each of the classes is contained within the superordinate class (i.e., ‘animals’) and that therefore each of the individuals is also contained within the superordinate class (Winer, 1980). The final step involves subtracting the relevant subclass (i.e., ‘dogs’) from the superordinate class (‘animals’), which is considered proof of the ability to perform asymmetric class containment, as doing so involves understanding that the superordinate class incorporates the subordinate classes within its own category and not vice versa (Winer, 1980). Apart from these individual skills, it is also maintained that combining the repertoires appropriately is required, and thus children may also fail due to centration, which is the inability to focus on more than one part of a problem at a time (Hendricks & Pasnak, 1999). It is argued that persistent exposure and practice with class inclusion tasks results in the maturation of a child's schematic representations, eventually enabling them to pass the test.

While Piaget claimed that class inclusion responding first emerged between the ages of 7 and 8 (Inhelder & Piaget, 1964), other researchers have disputed this. For example, Winer's (1980) review suggested that children might only become capable of class inclusion by age ten or later (Ahr & Youniss, 1970; Billow, 1975; Carson & Abrahamson, 1976; Hooper, Sipple, Goldman & Swinton, 1979; Kofsky, 1966; Lovell, Mitchell, & Everett, 1962; Meadows, 1977). Winer (1980) also argued that there might be a number of methodological limitations of the class inclusion task that might facilitate responding, including verbal facilitation, or cues highlighting the superordinate category. Winer proposed that these limitations could be responsible for variation in relation to the precise age at which class inclusion seemed to emerge.

Previous research has also evaluated the potential efficacy of class inclusion training for children under the age of 7 (e.g., Goswami & Pauen, 2005; Pasnak, Cooke & Hendricks, 2006; Siegler & Svetina, 2006). Studies have employed various techniques in this context including contingent verbal feedback (Brainerd, 1974; McCabe & Siegel, 1987), induced counting (Gash, 1982; Judd & Mervis, 1979; McCabe & Siegel, 1987), analogical reasoning (Goswami & Pauen, 2005), and asking participants to provide justifications for their responses (Siegler & Svetina, 2006; Thomas & Horton, 1997; Wohlhill, 1968). Such research has indicated that while it seems possible to train class inclusion in children as young as four, there remains a lack of data regarding the generalisation and maintenance of the results. For example, while McCabe and Siegel's class inclusion training was initially successful with children aged 4 to 6, when they tested for maintenance of training effects three months later, they found that the effects had dissipated.

Qualitative Class Inference Tasks. Markman (1989) suggested that in order to more properly and precisely assess the emergence of hierarchical classification, it was necessary to assess the emergence of relevant core properties of hierarchical classification separately from each other. Consistent with this approach, a number of researchers have developed qualitative class inference tasks in order to ascertain the precise age at which properties such as transitive class containment and asymmetric class containment develop (Deneault & Ricard, 2005, 2006; Greene, 1994; Johnson, et al., 1997; Smith, 1979).

Qualitative class inference tasks assess these properties by presenting children with quantified questions about inclusion (e.g., "Are all Xs Ys?"), and deductive qualitative inference questions as to whether a more or less inclusive category possesses a particular property or characteristic (e.g., "Xs have _____. Do all Ys have _____?"). When assessing transitivity inferences, individuals might be presented with the following statement and question, "A dax is a cat. Is a dax an animal?" (Deneault & Ricard, 2006). Asymmetric class containment might be assessed using a statement and question such as: "A dem is an animal. Is a dem a zebra?" (Deneault & Ricard, 2006). By evaluating the responses to such questions, researchers have tried to gauge the precise age at which transitivity and asymmetry comprehension emerge, and, amongst other questions, the sequence in which they occur.

Smith (1979) was one of the first researchers to employ assessment tasks such as those just discussed and thus she is frequently cited within the extant literature. She used a deductive task in which four-year-old children had to form qualitative inferences about class inclusion (see Blewitt, 1989, 1994; Deneault & Ricard, 2006; Markman, 1989). However, this study had a number of methodological limitations and also presented an incomplete data analysis. For example, order effects in terms of question presentation may have been an issue, partly as a result of an overly lengthy questioning period. Furthermore, Smith failed to compare participants' performances in relation to transitive and asymmetric class containment questions (Deneault & Ricard, 2006).

Johnson and colleagues (1997) replicated Smith's (1979) experiment by presenting 3-, 5-, and 7-year-olds with quantified questions about inclusion (e.g., "Are all X's Y's?") and deductive qualitative inference questions asking if a more or less inclusive category had a specific characteristic (e.g., "X's have _____. Do all Y's have ____?"). Findings indicated that all age groups performed similarly in relation to both quantified and qualitative questions, with all age groups providing accurate responses to half of the included questions, ultimately failing to demonstrate adequate transitivity or asymmetry comprehension and showing no differences in either transitive or asymmetrical responding across age groups. While Markman (1989) initially maintained that transitivity and asymmetry comprehension should precede class inclusion understanding, given the lack of variability in responding between age groups this result suggests that qualitative inference questions may not be as simple as originally anticipated by Markman (1989) and Blewitt (1989). However, researchers have indicated that these results might have occurred as a result of the participants' previous familiarity with the categories employed within the task (Deneault & Ricard, 2006). For example, in Johnson et al.'s task, children were able to answer half of the questions by using their knowledge of the categories (e.g., "All fire trucks have sirens. Do all trucks have sirens?") (Johnson, et al., 1997). Furthermore, the researchers also failed to analyse transitive and asymmetric class containment scores separately.

More recently, Deneault and Ricard (2006) employed qualitative inferential question tasks to assess transitive class containment and asymmetric class containment in a sample of children aged five, seven, and nine years. They found that while five-year-olds exhibit reasonable knowledge regarding the transitivity of inclusion relations, only the seven- and nine-year olds seemed able to reliably perform transitive class containment. Other research has found that the development of a robust understanding of asymmetry emerges comparatively slowly, and generally does not reach completion until the age of nine (Greene, 1994).

While the qualitative inference task has a number of benefits not possessed by the class inclusion task, Deneault and Ricard (2006) have highlighted the need for a more concrete definition of an acceptable asymmetric class containment response. One other criticism is that some of the questions posed are 'indeterminate' in nature, meaning that a definitive answer cannot be given. For example, while questions assessing transitivity are determinate, meaning that there is a definite answer to the question (e.g., the answer to the question "A dax is a cat. Is a dax an animal?" is 'Yes'), questions relating to asymmetry are indeterminate, meaning that the answer is unknowable (e.g., "A dem is an animal. Is a dem a zebra?"). Given the fact that before the age of 9- to 10-years-old children show significant difficulty identifying conditions in which there may be no conclusions (Byrnes & Overton, 1986; Deneault & Ricard, 2006; Horobin & Acredolo, 1989; Piérault-Le Bonniec, 1980; Pillow, Hill, Boyce, & Stein, 2000), this has been argued to pose a methodological limitation of the qualitative class inference task. This may also provide an explanation for children's poor asymmetry performance on class inference questions when compared to their transitivity performance.

Johnson and colleagues (1997) have suggested that an alternative to testing for asymmetry via qualitative inference might be to instead test for unilateral property induction between stimuli (i.e., a category induction test). Johnson maintains that children who successfully display a transfer of properties consistent with this relation may be presumed to understand the features of both transitive class containment and asymmetric class containment. However, it can be argued that this contradicts the idea that the three properties of hierarchical classification (i.e., transitive class containment, asymmetric class containment and unilateral induction) are independent of each other. Nevertheless, the use of category induction to test hierarchical categorisation has become more popular since. Category induction tasks have since been used to examine categorisation responses among individuals with autism (Soulières, Mottron, Giguère & Larochelle, 2011), to help assess the 'spacing effect' in children's learning and memory (Vlach, Sandhofer & Kornell, 2008), and to determine children's sensitivity to speaker intent in category induction (Jaswal, 2004).

Recent research. More recent cognitive developmental research into hierarchical classification has occurred in the context of cognitive developmental work assessing classification repertoires more generally across age groups ranging from early childhood to adulthood (Deng & Sloutsky, 2015; Rabi, Miles & Minda, 2015; Rabi & Minda, 2014). This work has examined classification strategies (e.g., rule-based category learning, similarity-based category learning) employed across different age groups and the impact of specific training (e.g., classification training and inference training) on category representation, ultimately focusing on the question as to whether and how the ability to categorise might change over the course of development. For example, Rabi and Minda (2014) examined the categorisation performances of adults as well as young children between the ages of 4 and 11 to determine developmental trends in categorisation responding. The results indicated that categorisation performance improved with age, with younger children demonstrating the greatest rule-based deficit when compared to older children and adults. The research conducted tracked changes in categorisation performances across childhood, providing information regarding the ages at which performance begins to improve and mirror that of adults.

Summary and Conclusion. Findings suggest that categorisation and hierarchical classification skills develops sequentially over time, though there remains debate regarding the specific age at which particular repertoires emerge and fully develop (see e.g., Johnson et al., 1997). Hierarchical classification initially emerges based on formal characteristics between the to-be-classified stimuli. Following this, children then learn to label the stimuli and learn of the properties and characteristics of categories at varying levels of generality (Blewitt, 1994; Johnson et al., 1997). Classification based on formal characteristics gradually shifts towards classification based on conceptual relations between classes (Johnson et al., 1997).

While cognitive developmental research has generated a number of useful procedures for measuring responding consistent with hierarchical classification, as has been discussed, this approach has several philosophical limitations including its invocation on the one hand of hypothetical (mentalistic) structures and processes (e.g., qualitative changes in schemata) and on the other of an inherent process of biological maturation to explain the development of hierarchical classification. Both these features mitigate against a strategy of specification of manipulable variables that might aid intervention in research or applied work. This is problematic for researchers or practitioners endeavouring to augment or facilitate hierarchical classification repertoires in young children.

In contrast to both the cognitivist and developmentalist influences underlying the work just reviewed, a key underlying philosophical assumption of behaviour analytic science is the importance of achieving not just theoretical prediction but also influence with respect to human behaviour. A core aim of the research program reported in the current thesis is to develop an effective behaviour analytic understanding of classification repertoires and in particular hierarchical classification in children. As such we turn now to review relevant behavioural research.

Behavioural Research

Classification has received considerable attention within behaviour analysis, perhaps not least because classification in itself is a core component of the behaviour analytic understanding of behaviour. For example, in relation to the operant, the stimulus and its subsequent response classes are crucial in this relationship, such that stimulus classes exert considerable influence over behaviours (Fields & Reeve, 2000; Zentall, Critchfield, & Galizio, 2002), and stimuli are subsequently considered to belong to the same class when a shared set of responses are emitted in their presence by an organism (Barnes-Holmes, Hayes, Dymond & O'Hora, 2001; Keller & Schoenfeld, 1950; Zentall, et al., 2002). The majority of behavioural research concerned with categorisation has centred on the formation of stimulus classes and researchers have been concerned with the development of three particular types of stimulus classes - perceptual classes, associative classes, and common object or natural language classes (Fields & Reeve, 2000).

Classes: Theory & Research. Perceptual classes as defined within behavioural psychology are similar to those described by perceptual classification work within developmental psychology, as both perspectives view perceptual classes as categories or groups of stimuli which share physical features (i.e., include perceptually similar stimuli; Fields & Reeve, 2000; Gelman & Meyer, 2011). Behavioural psychologists have examined perceptual classes among both human and non-human populations and have generally modelled perceptual classes using protocols that establish and assess simple discriminations and primary stimulus generalisation (Fields, Reeve, Adams, Brown & Verhave, 1997).

Associative classes contain members that share no formal features, but are related based on common functional properties (Zentall et al., 2002). One domain of research that has contributed substantively to our knowledge in this respect is work on stimulus equivalence (Galizio, Stewart & Pilgrim, 2001, 2004; Lane, Clow, Innis & Critchfield, 1998; Sidman, 1971). Matching-to-sample (MTS) protocols are typically employed in order to establish equivalence relations between physically dissimilar stimuli. Participants are first trained in a sequence of interrelated conditional discriminations with both sample and comparison stimuli, and are subsequently probed for the emergence of derived equivalence relations. For example, participants might be taught to select arbitrary A stimuli with arbitrary B stimuli (e.g., A1 – B1, A2 – B2) and to select the arbitrary B stimuli with arbitrary C stimuli (e.g., B1 – C1, B2 – C2). They may then show several untaught or derived relations including derived symmetrical responding or reversal of the taught relations (e.g., B1 – A1, B2 – A2, C1 – B1, C2 – B2) as well as derived transitive and symmetrical transitive responding based on the combination of taught relations (e.g., A1 – C1, A2 – C2, C1 – A1, C2 – A2). Sidman (1971), who provided the first empirical demonstration of this phenomenon, suggested that the emergence of this pattern meant the stimuli were now mutually substitutable or equivalent and indeed, further work, such as that on transfer of function (e.g., Catania, Horne & Lowe, 1989; Wulfert & Hayes, 1988) whereby stimuli in derived equivalence relations acquire the functions of other stimuli in that relation, bear that out.

The third type of stimulus class are natural language categories, which are seen as relating to real life, multifaceted and naturally occurring everyday classes (Fields & Reeve, 2000, p. 74). In a natural language class, all members may be substituted for one another, ultimately coming to function interchangeably (as in the case of the other two class types), but only some members share features (e.g., actual animals, the vocal verbal stimulus *'animal'*, and the textual stimulus "animal"). Behaviour analysis has typically conceptualised natural language categories in terms of generalised equivalence classes (Adams, Fields & Verhave, 1993; Fields, Adams, Buffington, Yang & Verhave, 1996; Fields & Moss, 2008; Galizio et al., 2004; Rehfeldt, 2003), which contain stimuli that are both perceptually and arbitrarily related (Fields & Reeve, 2000). For example, the aforementioned class of animals is a generalised equivalence class, comprised of textual stimuli, including “cat”, “dog”, and “sheep”, in addition to the sounds that correspond to those textual stimuli and the actual stimuli themselves with which the textual and auditory stimuli are related.

The foregoing has provided a brief introduction to the conceptualisation and investigation of classes and classification within behaviour analysis. This suggests that there has been a substantive quantity of work in this area, much of which has converged on the conception that both physical properties and abstract relations play an important role in classification responding, and much of which has employed derived equivalence relations to model the abstract relations involved. The foregoing is useful background, and will inform the approach taken. However, the core focus of the current thesis is hierarchical classification rather than just classification per se. Thus for the current purposes we will next begin to examine recent behaviour analytic work on hierarchical classification more specifically. From the present point of view, Relational Frame Theory (RFT) provides a potentially useful approach for modelling and investigating hierarchical classification and thus significant space will be spent introducing RFT and explaining the basis of the RFT approach to hierarchical classification. However, this will be the focus of the next chapter. In what follows, in the remainder of the present chapter, we will consider behaviour analytic studies of hierarchical classification that preceded the RFT approach.

Hierarchical classification. Griffee and Dougher (2002) conducted the first behaviour analytic study specifically designed to model hierarchical classification and establish responding consistent with a natural language hierarchy. They attempted to model the conditions required to establish such responding by using stimuli with shared physical characteristics and stimulus functions.

An example of one such natural language hierarchy is the taxonomy "*Labrador* is a member of the class *dog*; *Dog* is a member of the class *mammal*". People respond to stimuli in line with such hierarchical groups under the control of a mixture of relevant functions, physical features and context. For instance, an individual might own two labradors and respond differently to each as he can easily distinguish between them based on physical features. However, people who have not learned these features would find it difficult to distinguish between them. In contrast, they would likely be able to distinguish between a labrador and a bulldog as they differ far more significantly from one another than the two labradors. As a result of these greater physical differences, the two separate breeds elicit different responses in a greater variety of contexts than a pair of labradors would. Despite these differences, there are also several contexts in which a labrador and a bulldog might be treated similarly, however. For example, when addressing their daily needs, a person might respond to them both as dogs, while in discussing particular behavioural and physical similarities with cats, an individual might refer to them as mammals.

To empirically model features of responding in accordance with natural language hierarchical classes such as these, Griffee and Dougher (2002) established contextual control over responding to four similar (triangle) stimuli that differed along a physical continuum with respect to one feature (i.e., length of base). These stimuli represented objects or events that might be placed in a taxonomy such as that just discussed. In addition, colour background was manipulated to represent a range of different contexts: superordinate, intermediate, and subordinate. Across all contexts, participants received points (exchangeable afterwards for money) as reinforcement for correct responding and received more points the more accurate their responding.

The green background modelled a superordinate context. Participants were exposed to four different-sized triangles as samples and seven nonsense syllables (which might be coded H1, H1.1, H1.2, H1.1.1, H1.1.2, H1.2.1, H1.2.2) as comparisons. They were trained to choose the same nonsense syllable (H1) from the comparison array in the presence of each of the four triangles. This modelled responding to a broad class of stimuli (e.g., responding to all members of an array of different animals as mammals). The red background modelled an intermediate context. Participants were trained to choose two other nonsense syllables (H1.1 and H1.2) from the same seven-syllable comparison array. Choices of H1.1 were reinforced in the presence of the two smaller-base triangles and choices of H1.2 were reinforced in the presence of the two larger-base triangles. This was analogous to responding to a stimulus using an intermediate class term (e.g., dog, cat) in a context in which that is appropriate (e.g., talking about what kind of pets I own). Finally, the yellow background modelled a subordinate context. Participants were trained to choose a particular nonsense syllable from the four remaining nonsense syllables (i.e., H1.1.1, H1.1.2, H1.2.1 and H1.2.2) in the presence of each of the four triangles respectively. This modelled responding to smaller (more specific) subclasses (e.g., responding differently to each of two different breeds of dog and two different breeds of cat). After this pattern of responding was established, the participants were then tested with novel triangle stimuli that differed from the originals in terms of base length, and they duly showed generalisation of the response pattern. Next, in order to simulate a response function other than naming the stimuli, the participants were trained to press onscreen buttons under contextual control and generalisation was again tested. Finally, to introduce arbitrary (language-like) relations into the model, derived relations between the nonsense syllables and buttons and the nonsense syllables and triangles were tested.

Griffee and Dougher (2002) thus demonstrated responding analogous in certain respects to hierarchical classification. Differential responding to stimulus features was brought under contextual control and the pattern of control appeared hierarchical. Responding in the superordinate context was controlled by a common feature and stimuli shared a common stimulus function; however, in subordinate contexts, responding was based on more differentiated features and stimuli shared more specific stimulus functions. This study thus represented an initial step towards a behaviour analytic conceptualisation of hierarchical classification. However, this model was only preliminary. Of particular importance, there was no test of any of the three features that characterise hierarchical classification according to mainstream psychology, namely transitive class containment, asymmetrical class containment, and unilateral property induction.

Noting this omission from Griffee and Dougher (2002), Slattery, Stewart & O'Hora (2011) sought to replicate and extended this previous model of hierarchical classification by including a test for one of the three key features, namely transitive class containment, and remediation training should responding in accordance with this feature fail to emerge. Experiment 1 replicated a number of the main components of Griffee and Dougher's (2002) protocol but differed from it by testing transitive class containment by employing stimuli that were physically dissimilar to those employed during training. The findings were that only two of the five participants demonstrated the required pattern of responding. Experiment 2 examined whether repeated exposure to the protocol might result in correct responding but in this case none of the three new participants responded correctly. The final experiment involved replacing the stimuli employed in the first two experiments by arbitrary stimuli. This experiment proved relatively successful, with all three new participants exhibiting responding consistent with transitive class containment.

While Slattery et al. (2011) marked an advance in modelling aspects of hierarchical classification, the researchers themselves called attention to a number of important points. First, it was unclear whether the effect observed within the study was transitive class containment; they argued that it may have been a functionally simpler pattern of behaviour. Second, even if transitive class containment was indeed shown by the participants the question remained as to how the adapted Griffee and Dougher protocol worked to facilitate this result. They suggest that is possible that the inclusion of arbitrary stimuli might have cued a pattern of abstract hierarchical relational responding including transitive class containment. As they pointed out, while physically similar objects are crucial to the initial foundation of a hierarchical classification repertoire, the hallmark of a completely established repertoire of hierarchical classification is the capacity to respond correctly to novel classes that are independent of physical relations between stimuli (this suggested developmental progression is the same as that posited within mainstream theory). In any event however, it was uncertain in the case of this particular study whether abstract stimuli had influenced the production of transitivity and indeed, even whether transitivity had been definitely been shown. As Slattery et al. (2011) also reflected, even if the study had indeed effectively demonstrated transitive class containment, it did not show either of the other two features that earlier theorists have maintained are part of hierarchical classification, namely asymmetry or unidirectional property inheritance. Hence, they argued, additional work was required to provide a more adequate model of hierarchical classification and they recommended Relational Frame Theory (RFT) as the theoretical basis for such a model.

Summary and conclusion. To date, behavioural research has investigated and, in a number of cases, successfully modelled several phenomena involved in classification and hierarchical classification as previously discussed under the rubric of cognitive and developmental theory and research. These include, for example, perceptual (formally based) and conceptual (abstract) classes, associative classes, and transitive class containment. In doing so, this research has contributed substantively towards creating a functional analytic understanding of the processes underlying classification and hierarchical classification. This is an important part of the background to the research reported herein. However, as suggested in the previous section, there are several aspects of hierarchical classification in particular that pre-RFT research did not investigate or model. Slattery et al. (2011) recommended that RFT could be used to model this phenomenon more successfully than had been seen in previous work. Since then several studies have used RFT to model hierarchical classification in adults and this work provides a key element of the background to the research presented herein which used RFT to assess and train hierarchical classification responding in children. As such, in the next chapter, we will introduce and discuss RFT as well as RFT-based work on hierarchical classification in adults.

Chapter 2

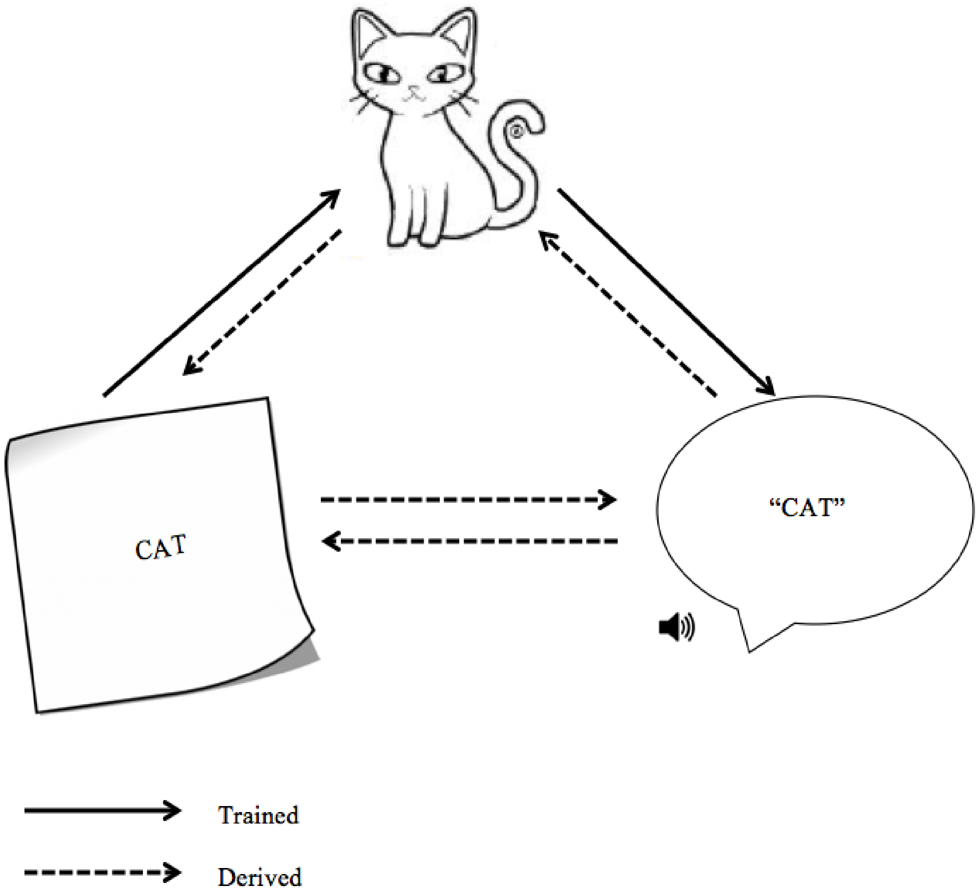
Relational Frame Theory & Hierarchical Classification

As previously indicated, the research in this thesis focuses on the assessment and training of relational repertoires theorised to form the basis of classification responding in young children. The theoretical basis of this approach is Relational Frame Theory (RFT) which considers language and cognition in terms of derived or arbitrarily applicable relational responding. As such, the current chapter will focus on introducing this approach starting with early work on derived relational responding and then proceeding to outline the theoretical and empirical basis for RFT. Following this, the RFT approach to hierarchical classification will be outlined. Finally, the chapter will provide a brief overview of the empirical work reported in the ensuing chapters of the thesis.

Derived Relational Responding: Stimulus Equivalence

The most well-known and well-investigated form of derived relational responding is stimulus equivalence. This was first demonstrated in Sidman's seminal (1971) study, the original aim of which was to teach reading skills to an intellectually disabled adolescent. Prior to the intervention, the participant had never demonstrated any reading comprehension but had shown evidence of a matching repertoire; specifically, he could select pictures [A] given the corresponding spoken words [B]. In addition, Sidman used a match-to-sample (MTS) protocol to teach him to select textual stimuli (printed words) [C] in the presence of spoken words [B]. Thus, at that point, he had learned picture [A] – spoken word [B] relations and text [C] - spoken word [B] relations. Subsequently, without any additional training, he showed several further untaught or *derived* performances including for example, matching printed words [C] to pictures [A], and pictures [A] to spoken words [C]. Sidman suggested that the pattern shown was such that he seemed to be treating the A, B and C stimuli as if they were mutually substitutable or equivalent and hence he called this type of performance 'stimulus equivalence'.

Figure 2.1. Representation of Stimulus Equivalence.



The core of the stimulus equivalence effect was that training involving a series of related conditional discriminations could result in untrained or derived relations in accordance with a pattern of derived equivalence. Sidman and colleagues replicated and investigated this phenomenon a number of times, and based on his investigations, Sidman suggested that derived equivalence involved three key emergent properties which are present following such initial conditional discrimination training, namely, reflexivity, symmetry, and transitivity (Sidman, & Tailby, 1982).

Reflexivity is assessed by testing for generalised identity, which involves matching a single stimulus to itself (e.g., matching a picture of a dog to another picture of a dog). Symmetry is assessed by testing for reversibility, which follows training to establish the selection of a comparison A (e.g., picture of a dog) upon sample B (e.g., the text “dog”), or A-B matching. Once this relation has been established, symmetry is then assessed by testing for the reversal B-A matching. For example, when presented with sample B (the text “dog”) an individual will then select the comparison stimulus A (picture of a dog) from an array of comparisons. Transitivity requires a minimum of three stimuli and involves the demonstration of a novel conditional discriminative performance based on the training of two related conditional discriminative performances. For example, if an individual learns to match stimulus B (the text “dog”) in the presence of A (picture of a dog), and also to match C (the auditory stimulus “dog”) in the presence of A, then transitivity is observed if they demonstrate an untrained A-C match (i.e., matching the picture of a dog to the auditory stimulus “dog”) or the reversal of this (i.e., C-A; in this case, matching the auditory stimulus “dog” to the picture of a dog). If all three properties (i.e., reflexivity, symmetry and transitivity) are observed, then an equivalence class or equivalence relation has been successfully established.

Since Sidman (1971), numerous other studies have investigated the phenomenon of derived equivalence. It was the subject of intense study for a variety of reasons. First, it seemed to contradict traditional conceptions. It couldn’t be explained in terms of traditionally known behavioural principles including either classical or operant conditioning or generalisation (see e.g., Barnes, 1994; Hughes & Barnes-Holmes, 2016). Second, it seemed to provide a very useful paradigm for examining and boosting generative performance. For instance, researchers have employed stimulus equivalence to model generativity in grammar (e.g., Wulfert & Hayes, 1988) while Sidman (1971) and many other studies since have shown its potential utility in the applied setting where generative performance is hugely important. Such studies have often employed the transfer of function effect wherein a psychological function trained to one stimulus in a derived equivalence relation then appears in other stimuli in the relation without further training (e.g., Murphy, Barnes- Holmes & Barnes-Holmes, 2005; Rehfeldt & Root, 2005). For example, Rehfeldt and Root (2005) examined transfer of function through derived equivalence relations with three adults with severe developmental disabilities. Participants were first taught to mand for preferred items using picture (i.e., Picture Exchange Communication System; PECS). The researchers used a stimulus equivalence paradigm to establish conditional discriminations via a matching-to-sample procedure between pictures of preferred items and their dictated names (i.e., A-B) and between the dictated names and their corresponding text (i.e., B-C). As a specific behavioural function was established for A (i.e., exchanging pictures of preferred items for access to these items) that function then transferred to the indirectly related C stimulus (i.e., exchanging the corresponding text for these items) without explicit training, thereby resulting in the transfer of function through derived equivalence relation. For example, a child who learns to mand for Coke and then learns that "cola" is another word for Coke may subsequently mand for "cola" without a direct history of reinforcement for doing so (see also Murphy & Barnes-Holmes, 2009).

Third, and most importantly, equivalence was linked theoretically and empirically with human language (e.g., Stewart & Roche, 2013). Since Sidman (1971) many studies have indicated that stimulus equivalence and language are closely related (e.g., Barnes, McCullagh, & Keenan, 1990; Ogawa, Yamazaki, Ueno, Cheng & Iriki, 2009). Sidman's results themselves indicated that the stimuli had acted as mutually substitutable "symbols" despite having no previous history of pairing or any direct relationship. Since Sidman (1971), extensive research has indicated that equivalence can be obtained with various stimuli, populations and procedures (for a review see Sidman, 1994; 2000; 2009); however, it can only be achieved by verbally-able humans and has not been demonstrated in verbally-disabled humans or non-humans (see Barnes, 1994; Barnes, McCullagh, & Keenan, 1990; Brino, Campos, Galvào & McIlvane, 2014; Devany, Hayes, & Nelson, 1986; Dugdale & Lowe, 1990, 2000; Lionello-Denolf, 2009). Additional research has also indicated that young children who initially fail an equivalence test subsequently demonstrate successful equivalence responding once trained to tact the stimuli involved (Dugdale & Lowe, 1990; Eikeseth & Smith, 1992). Research has shown that equivalence training can remediate language deficits in verbally-disabled people (see Barnes, 1994; Cowley, Green & Braunling-McMorrow, 1992; Matos & d'Oliveira, 1992). In addition to this, the previously discussed transfer of function effect offers an analogue of linguistic control (e.g., Dougher, Auguston, Markman, Greenway, & Wulfert, 1994; Murphy, Barnes-Holmes, & Barnes-Holmes, 2005; Rehfeldt & Root, 2005; Stewart & Roche, 2013).

Thus stimulus equivalence has inspired much excitement and research within the behaviour analytic community. At the same time though, stimulus equivalence is an empirical phenomenon, not a theory. As already outlined, it was not something that would be predicted based on traditional principles. As such, it required explanation. In the 1990's a number of theories emerged to explain it. These included (a) the theory that stimulus equivalence relations arose from the reinforcement contingency itself (Sidman, 2000); (b) the suggestion that multi-nodal derivative relations begin to acquire control after one-node derivative relations have begun to exert stimulus control (see Fields, Adams, Verhave & Newman, 1990; Lazar, Davis-Lang & Sanchez, 1984; Saunders, Wachter, & Spradlin, 1988), (c) naming theory which suggests that the "*name relation*" is the basic unit of verbal behaviour and is learned and comes to symbolise objects and events in the real world (Horne & Lowe, 1996), and (d) Relational Frame Theory (Hayes, Barnes-Holmes & Roche, 2001). Of these theories, arguably the most adequate and unquestionably by now the most empirically well supported is Relational Frame Theory (e.g., Stewart & Roche, 2013). RFT suggests that derived equivalence is just one pattern of derived relational responding and that in fact there are multiple other such patterns. In addition, it explains that the phenomenon of derived relational responding is a learned (operant) behaviour that enables human language and that this is the reason for the empirical link between derived relations and language. RFT is the approach taken to explaining classification in the current thesis and thus at this point, I will introduce and describe this theory.

Relational Frame Theory

As suggested, RFT sees derived relational responding as learned operant behaviour. It explains it as generalised contextually controlled relational responding referred to as arbitrarily applicable relational responding (AARR), also known as relational framing. This concept will be explained in what follows.

Relating is a generalised pattern of behaviour, which involves responding to at least one stimulus in terms of at least one other stimulus (Hughes & Barnes-Holmes, 2016). Both human and non-human populations can respond relationally to stimuli and events; however only humans are capable of a more complex variety of relating behaviour (i.e., AARR), one which profoundly affects their interaction with their environment. In order to explain this more fully I will examine the two varieties of relating - non-arbitrarily applicable and arbitrarily applicable.

Non-arbitrarily applicable relational responding (NAARR). Non-human populations including mammals, birds, fish, and insects can be taught to respond to the relations between and among stimuli in the environment (see Frank & Wasserman, 2005; Harmon, Strong, & Pasnak, 1982; Yamazaki, Saiki, Inada, Iriki, Watanbe, 2014). For such species, these relating responses are characterised by two core features. First, they are based on a previous learning history involving direct experience with the stimuli. Second, they are based on the physical characteristics of the stimuli (Giurfa, Zhang, Jenett, Menzel, & Zrinivasan, 2001; Harmon et al., 1982; Hughes & Barnes-Holmes, 2016). In RFT, this pattern of responding is called non-arbitrarily applicable relational responding (i.e., NAARR) because responding is based on the formal or physical properties (e.g., size, shape, colour, quantity, etc.) of the stimuli (e.g., Stewart & Roche, 2013; Hughes & Barnes-Holmes, 2016).

Consider a pigeon exposed to a learning task in which a sample stimulus (e.g., a red circle) appears at the top of a computer screen while two comparison stimuli (a red and a green circle) are located at the bottom of the screen. For trials in which the red circle is presented as sample, selecting the red circle comparison will be reinforced, while on trials in which the green circle is the sample, selecting the green circle comparison will be reinforced. The pigeon is exposed to further training with stimuli of differing colours and shapes and is ultimately presented with a number of novel stimuli (that were never directly reinforced in the past) once they consistently respond correctly across a large number of trials. Research in this arena suggests that the pigeon will subsequently choose the shape at the bottom of the screen that is physically identical to the shape at the top of the screen, even when that particular response was never previously reinforced (Frank & Wasserman, 2005). This is an example of generalised relational responding but the key feature is that it is based on the physical properties (in this case colour) of the stimuli involved.

Arbitrarily applicable relational responding (AARR). While both human and non-human populations demonstrate NAARR, humans exposed to particular contingencies by the socio-verbal community also develop a more advanced type of relating, namely arbitrarily applicable relational responding (AARR) or relational framing, which involves relating stimuli primarily under contextual control and thus independent of their physical properties (Hughes & Barnes-Holmes, 2016). For example, if I tell even a relatively young child that X is more than Y then they will derive that Y is less than X; this is not based on physical properties but on the cues ‘more’ and ‘less’ which cue the relational frame of comparison. Furthermore, RFT maintains that this type of relational repertoire is an operant, one key context for the early acquisition of which involves learning non-arbitrary relations under contextual control (e.g., choosing the physically bigger or smaller of two objects in the presence of the cues ‘more’ and ‘less’ respectively). With sufficient exemplars of non-arbitrary relations, however, it is posited by RFT that this responding can come under the control of the cues alone and is then applicable in contexts without non-arbitrary relational support (e.g., if told that X is more than Y, a child can derive that Y is less than X in the absence of any non-arbitrary or formal differences between the relata). This is called arbitrarily applicable relational responding (AARR) because the relational response can be applied in any circumstance no matter what the actual physical relation between the relata (Hayes, Fox, Gifford, Wilson, Barnes-Holmes & Healy, 2001; O’Toole, Barnes-Holmes, Murphy, O’Connor & Barnes-Holmes, 2009).

Over the last two decades, RFT researchers have shown that patterns of AARR (i.e., relational frames) come in a variety of forms including, for example, coordination (e.g., “same as”; Cassidy, Roche & Hayes, 2011) which is the basis for equivalence responding, comparison (e.g., “more than/less than”; Berens & Hayes, 2007), opposition (e.g., “opposite to”; Cassidy, Roche, Colbert, Stewart & Grey, 2016), distinction (e.g., “different to”; Rehfeldt & Barnes-Holmes, 2009), analogy (e.g., A : B :: C : D, apple is to orange, as dog is to sheep; Stewart, Barnes-Holmes, Roche & Smeets, 2002), deixis (e.g., “I-you”, “here-there”, and “now-then”; McHugh, Barnes-Holmes, Barnes-Holmes, Whelan & Stewart, 2007), temporality (e.g., “before/after”; O’Hora, Peláez, Barnes-Holmes, Rae, Robinson & Chaudhary, 2008), and hierarchy (e.g., “type of”; Slattery & Stewart, 2014). RFT researchers have also offered evidence that (a) people respond consistent with these frames and (b) they can be trained up when weak or absent (e.g., Barnes-Holmes, Barnes-Holmes & Smeets, 2004; Barnes-Holmes, Barnes-Holmes, Smeets, Strand & Friman, 2004; Berens & Hayes, 2007; Carpentier, Smeets & Barnes-Holmes, 2003; McHugh, Barnes-Holmes, Barnes-Holmes & Stewart, 2006; Roche & Barnes, 1997; Rosales, Rehfeldt & Lovett, 2011; Steele & Hayes, 1991). Despite the variation in patterns of framing, three fundamental properties are present in all cases: mutual entailment, combinatorial entailment, and transformation of function.

Mutual entailment refers to the feature of relational framing by which a unidirectional relation from stimulus A to stimulus B in a specific context entails a second unidirectional relation from B to A. For example, if a person is told that arbitrary stimulus A is a member of arbitrary stimulus B, then they can derive that B contains A. Combinatorial entailment is the phenomenon whereby two stimulus relations are combined to derive a third relation. For example, if an individual is told that A is a member of B, and B is a member of C, then they may derive that A is a member of C and that C contains A. The transformation of functions is crucial from an RFT perspective as it is the means by which language can influence our behaviour. If two arbitrary stimuli A and B partake in a derived stimulus relation and stimulus A has a psychological function, then under specific conditions the stimulus functions of B may be transformed in accordance with that relation. For example, imagine that, after deriving that C is a class that contains both B and A, a person is then informed that all C’s can be eaten. If they are then asked whether A and B can be eaten, even without seeing A or B and without any additional training or instruction, they may reply that they can be. In this way, the functions of A and B are transformed via the derived relation with C (see e.g., Barnes-Holmes, Barnes-Holmes, & McHugh, 2004).

AARR and Language. From the RFT perspective, the most important thing about arbitrarily applicable relational responding is that it is believed to underlie language and cognition (e.g., Dymond & Roche, 2013; Hughes & Barnes-Holmes, 2016). As discussed earlier in relation to equivalence, there is substantial evidence that the capacity to derive relations in accordance with equivalence (i.e., coordinate frames) is similar to language. For instance, spoken and written words bear little to no physical similarity to their referents, however people respond to these stimuli as though they are the same (e.g., yelling “bomb” on a bus may elicit the same fear response as actually seeing a bomb on a bus). In addition to this, the ability to derive any number of complex relations between stimuli, despite having limited exposure to them is also reminiscent of the generativity of language (Hughes & Barnes-Holmes, 2016). AARR, as demonstrated in empirical research, bears a striking resemblance to language skills. As such, this has lead researchers to believe that it is a crucial process of language (Törneke, 2010).

The ability to derive complex relational networks when exposed to minimal explicit training may account for the complexity of human language (Barnes-Holmes, Hayes, Dymond & O’Hora, 2001). RFT research has indicated that such AARR may only be performed by language-able humans. For example, Devany and colleagues (1986) assessed as to whether language abilities were related to an individual’s ability to establish equivalence classes by comparing the performance of three groups of children (typically developing children, developmentally disabled children with speech, and developmentally disabled children with a language deficiency). The results of this experiment indicated that language-able children demonstrated superior performances in relation to the stimulus equivalence tests than those children without language. These results indicated a positive correlation between language ability and stimulus equivalence performances. However, it can be argued that developmentally disabled populations may present with additional impediments, other than language abilities, so these results may not accurately reflect the relationship between language and AARR repertoires in such populations (McLay, Church & Sutherland, 2014). More recent empirical research has also indicated that AARR repertoires are correlated with cognitive and language abilities (Cassidy et al., 2016; Cassidy, Roche & Hayes, 2011; O'Hora et al., 2008).

Neuropsychological research has also indicated that AARR may be linked to language. This research has successfully demonstrated that when people are engaged in AARR tasks, the brain activity recorded during these tasks is similar to that observed during language-related tasks (Staunton & Barnes-Holmes, 2004; Dickins et al., 2001). For example, using an fMRI, Dickins and colleagues (2001) studied the brain activation of 12 participants who were engaged in MTS tests of AARR consistent with equivalence and a test of verbal fluency. The results indicated that the brain activation observed during the MTS tasks was similar to that observed during verbal fluency tasks. More specifically, both MTS and verbal fluency tasks activated dorsolateral prefrontal cortex and posterior parietal cortex bilaterally. Furthermore, in three of the four MTS tasks behavioural accuracy was significantly correlated with left lateralisation of dorsolateral prefrontal cortex activity. Dickins et al. (2001) provided empirical evidence that brain activation patterns during AARR tasks resemble those involved in the processing underlying language, without simultaneously activating regions of the brain associated with the simple sub-vocal articulation of stimulus names.

These findings support the RFT argument that AARR, the process of framing events or stimuli relationally according to the three criteria previously outlined (i.e., mutual entailment, combinatorial entailment, and transformation of function; Hayes et al., 2001; Törneke, 2010), could be a key process underlying language. Furthermore, AARR is an operant, and thus its development and potential training are specifiable in functional analytic terms, which is critically important in terms of training and practical intervention to boost linguistic and cognitive ability. In what follows, I will examine RFT research that has focused on training particular relational frames, both as an example of the core utility of this approach as well as background to the current doctoral work, which involves training frames relevant to classification.

It is believed by RFT researchers that AARR is typically learned by children as a result of their exposure to naturally occurring daily environmental contingencies that facilitate and establish these response patterns (e.g., Luciano, Gómez-Becarra, & Rodríguez-Valverde, 2007). In examining the acquisition of these relational repertoires, I will start by considering the relational frame of coordination, which is regarded as the first frame to emerge in early childhood (Lipkens, Hayes & Hayes, 1993; Luciano et al., 2007) and one which, as suggested before, is of fundamental importance for linguistic reference and thence language itself.

An infant might be taught to orient or point towards a particular stimulus when presented with a novel word, such as: “Grandma”, (Infant looks at

Grandma), (Hear Name A – Look at Person B). They might subsequently be taught to vocalise this person's name in their (i.e., the person's) presence: (Mother points to Grandma) "Who is this?" (Infant: "Grandma"), (See Person A – Vocalise Name B). RFT researchers maintain that these symmetrical relations initially need to be trained bi-directionally (i.e., A-B; B-A); however, following sufficient exemplars of such bi-directional training the infant will begin to acquire the operant response pattern of bi-directionality, such that in future, training in one direction with a novel word-object pair in the context of object-naming is sufficient to produce derived relation responding in the untrained direction (McHugh & Reed, 2008). As a result of the training involved, it is believed that these specific contextual cues (e.g., "is") are established for symmetrical responding. This is coordinate responding with two stimuli; with further exposure to socio-verbal interactions more complex patterns of co-ordinate responding involving three or more stimuli (i.e., equivalence) might be similarly acquired. Furthermore, MET may also successfully establish responding consistent with relational frames other than coordination (e.g., distinction, opposition, comparison). Thus, RFT conceptualises AARR as a generalised response class produced via reinforcement across multiple exemplars. Once this response class has successfully been established, any stimulus may subsequently partake in any relational frame which has been acquired, regardless of their physical form.

Empirical research has been generated which has focused on assessing and training various repertoires of AARR among both young neuro-typical (NT) children and individuals with developmental disabilities. For example, Ruiz-Sánchez and Montoya-Rodríguez (2014) conducted a literature search that examined the contributions provided by RFT with child and adolescent populations with and without developmental disabilities. This search indicated that a total of 58 articles published between 1986 and 2014 met these criteria, of which 35 employed NT participants and 23 included atypically developing populations. This indicated that to date, a variety of different relational framing repertoires have been successfully trained including coordination (Luciano et al., 2007; Walsh, Horgan, May, Dymond & Whelan, 2014), distinction (Newsome, Berens, Ghezzi, Aninao & Newsome, 2014), comparison (Barnes-Holmes et al., 2004), opposition (Cassidy et al., 2011), and perspective taking (Weil, Hayes, & Capurro, 2011).

As previously described, coordination is considered to be one of the first relational frames to emerge and typically develops in early childhood and language development. As such it is challenging to evaluate procedures designed to establish relational responding consistent with this frame. However, research targeting coordinate AARR exists. For example, Luciano and colleagues (2007) successfully established receptive symmetrical responding with a 15-month-old infant by employing MET in bi-directional object-sound/sound-object relations with multiple objects. In this study, an object (A) was presented to the infant, and was labelled (B) by the experimenter (i.e., A-B relations). Training was then introduced to the infant, whereby she was instructed to select the object from an array (i.e., B-A relations). Thereafter the infant showed receptive symmetrical responding with novel stimuli, such that she could correctly choose a particular item from an array (B-A) for a novel object that had been previously presented and labelled by the experimenter (A-B).

The relational frame of distinction, or difference, is concerned with responding based on the differences between stimuli based upon a specific dimension. Newsome and colleagues (2014) recently assessed the efficacy of MET on reading comprehension under contextual cues consistent with coordination and distinction. Five typically developing 9- to 12-year-olds with poor reading comprehension, but strong reading abilities, were participants. The first phase of training involved teaching them to accurately identify the categories, features and functions of a number of everyday stimuli. Following this phase they were trained in a number of relational tasks concerning relational frames of coordination and distinction. For example, they were asked to describe how certain stimuli (e.g., a taxi and a cat) are the same as, or different from one another. They were also asked about activities, for example, “How is swimming in the pool [different from/like] playing in the park?” Results showed improvements in participants’ ability to accurately respond to similar relational tasks with unfamiliar stimuli, in addition to marked improvements in reading comprehension as measured by standardised assessments.

Comparative framing involves responding to one stimulus in terms of a quantitative or qualitative relation along a particular dimension with another stimulus or event. In a comparative frame as in some others, the bi-directional relation between stimuli is not a symmetrical one. For example, mutually entailed comparative relational repertoires are said to have been successfully established if an individual correctly derives that “B is less than A” when provided with the information that “A is more than B”. Barnes-Holmes and colleagues (2004) offered the first empirical study to establish comparative AARR. In this study, three young typically developing children between 4 and 6 years were exposed to MET to establish comparative responding among abstract stimuli. In the typical task, the child was told the comparative relationship between stimuli (e.g., A > B

> C) and then asked to select the stimulus that would buy the most sweets. All three children failed baseline tests. However, following MET (in conjunction with non-arbitrary comparative training for one child), all three showed generalised comparative AARR. These findings have since been replicated and extended, indicating the feasibility and efficacy of MET procedures for training comparative AARR (Berens & Hayes, 2007; Gorham, Barnes-Holmes, Barnes-Holmes & Berens, 2009).

RFT maintains that deictic relational framing (i.e., responding correctly under the stimulus control of the contextual cues I-You, Here-There, Now-Then) is the basis for perspective-taking (McHugh, Barnes-Holmes, & Barnes-Holmes, 2004). Deictic framing is argued to emerge following a history in which the child must respond correctly in accordance with questions and statements involving deictic cues (e.g., “What will you do here?”, “What happened then?”). McHugh et al. (2004) found that performance on these relational repertoires correlates with age. Rehfeldt and colleagues (2007) employed MET to assess deictic framing in typically developing children and children with autism. The protocol employed examined three perspective-taking frames (i.e., I-You, Here-There, Now-Then) across three levels of complexity (simple, reversed, and double-reversed) within a conversation format between the researcher and the child. They found that the children with autism made more errors on particular deictic tasks than the typically developing children. The researchers also successfully trained the typically developing children in particular deictic relational repertoires. More recently Weil, Hayes and Capurro (2011) extended this work by establishing deictic relational framing in children with autism.

To summarise at this point, RFT argues that AARR is an operant repertoire that explains the phenomenon of derived relational responding and that underlies linguistic and cognitive performance in humans. Furthermore, a growing body of research has shown a multiplicity of patterns of relational framing (e.g., coordination, distinction, opposition, comparison, deictic) and demonstrated that these repertoires can be trained in children both with and without developmental disabilities. In the last chapter we examined approaches from mainstream (cognitive-developmental) psychology and behaviour analysis towards the important repertoire of hierarchical classification. In common with other complex repertoires such as perspective taking, this repertoire can be approached as particular patterns of relational framing, more specifically containment and hierarchical relational framing. Furthermore, recently, RFT studies have begun to model hierarchical classification on this basis. At the same time, this research has been conducted solely with adults and has not yet addressed hierarchical relational responding among children or individuals with developmental disabilities. This was a core aim of the current doctoral work. Thus, in order to provide a background to this work, in the final section of this chapter, I will first introduce RFT theory and research on hierarchical classification before subsequently providing an overview of the empirical research conducted in the current doctoral programme.

RFT and Hierarchical Classification

According to Barnes-Holmes and colleagues (2004) the verbal relational skills afforded by RFT form the foundation of a spectrum of cognitive abilities correlated with educational achievement including classification. From the RFT point of view, classification can be conceptualised as involving particular types of framing. Perhaps the most relevant frame in this regard is the so-called containment relational frame, which is theorised to develop prior to the more complex hierarchical relational framing, and is also believed to support the development of the latter (Hayes, Barnes-Holmes & Roche, 2001).

Containment relational framing, as with other relational frames, can be theorised to originate in more basic non-arbitrary relations. For instance, this relational frame may initially develop with mutual non-arbitrary containment relations. For example, a child will learn to relate objects as being physically inside other objects (e.g., 'food is in the fridge', 'my foot is in my shoe') and that the latter contain the former (e.g., 'the fridge contains food', 'my shoe contains my foot'). Such non-arbitrary relations may sometimes be a little more complex, involving the combination of relations between three or more stimuli (e.g., 'the carrot is in the vegetable drawer and the vegetable drawer is in the fridge', 'my foot is in my sock and my sock is in my shoe').

As with other frames, it is believed that with exposure to sufficient exemplars the contextual control (e.g., cues such as "in" and "contains") may come to exert control over the relational response pattern without the need for the presence of particular physical relations. For example, the child may come to respond appropriately in accordance with such contextual control without seeing the stimuli involved or may respond correctly to a question concerning a hypothetical scenario such as 'If the sheep is in the field, and the field is in Mayo, then where is the sheep?' This is arbitrary relational responding in accordance with containment. Ultimately, both non-arbitrary and arbitrary containment framing lay the foundation for arbitrary hierarchical framing in which the person relates elements to a larger collective or class under the control of cues such as the words “type of”, “includes”, “member of”, and “contains”.

To date, a number of studies have modelled hierarchical responding as hierarchical relational framing. The first of these was Gil, Luciano, Ruiz and Valdivia-Salas (2012). In this study, non-arbitrary relational training was used to establish contextual cues for both hierarchical (‘includes’, ‘belongs to’) as well as ‘same’ and ‘difference’ relations. In the case of the hierarchical cues, for example, participants were exposed to 2-D picture stimuli inside 2-D outline stimuli and were trained to choose the former in the presence of the ‘belongs to’ cue and to select the latter in the presence of the ‘includes’ cue. Participants were then exposed to a similar procedure in which contextual cues for responding consistent with “same” and “different” relations were established by teaching participants to choose stimuli which were physically identical to one another when presented with a specific arbitrary shape (i.e., this arbitrary shape acted as a contextual cue for “same”), while also choosing to select stimuli which were physically different from one another when presented with a specific arbitrary stimuli (i.e., this arbitrary shape acted as a contextual cue for “different”). The next stage involved teaching and assessing the participants on all four cues (i.e., “includes”, “belongs to”, “same” and “different”) in tasks using stimuli inverbal relations, which the researchers assumed to have been previously established with the participants. For example, when a participant was provided with the contextual cue for “includes”, the participant’s selection of the word “vowels” when shown the word “alphabet” was reinforced. Participants were subsequently taught three 4-member equivalence relations involving abstract shapes and nonsense syllables by employing a standard one-to-many matching-to-sample (MTS) procedure.

Following this phase of training participants were then assessd for the emergence of these equivalence relations. The experimenters then used the previously established contextual cues of “includes” and “belongs to” in order to establish a potential arbitrary hierarchical relational network. Within this experiment, some of the stimuli used within the previous experimental stage (i.e., the stimuli used to establish three 4-member equivalence classes) formed the bottom-most level of the hierarchy, while the middle and upper levels of this hierarchy included novel (i.e., untrained) abstract shapes. Using a stimulus pairing and MTS procedure, participants were then taught associations between specific stimuli in the arbitrary hierarchical relational network previously formed (i.e., the relational network comprised of bottom, middle and upper levels), in addition to also being taught words describing specific stimulus properties (e.g., “sweet” and “cold”). In the final experimental phase, participants were then tested for the expected transformation of functions of arbitrary stimuli via hierarchical relations that had not been explicitly connected with specific properties within previous experimental phases. As an example, participants were taught that stimulus ‘X.1’ had the property of being cold and that it contained a number of other arbitrary stimuli including A1 and B1. They were tested for whether a stimulus C1, previously derived as being in the same class as A1 and B1, was also cold and they showed this predicted transformation of functions. This study demonstrated a number of empirical and methodological innovations including the establishing of contextual cues for containment relations and the demonstration of a format in which responding in accordance with multiple stimulus-relations (same, different, belongs to, includes) was probed through requiring selection of contextual cues for particular frames.

More recently, Gil, Luciano, Ruiz and Valdivia-Salas (2014) extended this initial study by showing additional patterns of derived hierarchical relations and by providing an improved set of controls over participants' performance. Within this study, Gil and colleagues analysed the transformation of stimulus function via hierarchical relational framing by analysing the transformation of stimulus function via hierarchical relations with eight adults using a computer task comprised of five phases. The stimuli employed within Gil et al. (2012) were re-used within their 2014 experiment. Again, arbitrary symbols were trained to serve as relational cues (i.e., "includes", "belongs to", "same", and "different"). As with Gil et al. (2012), three 4-member equivalence classes were taught and assessed (A1-B1-C1-D1; A2-B2-C2-D2; A3-B3-C3-D3). The "includes" and "belongs to" relational cues were then used to establish inclusion relations between the lower levels of the hierarchy (i.e., A1/B1, A2/B2, and A3/B3) and the stimuli X.1, X.2 and Y.1 respectively. Following this, inclusion relations were then established between X.1/X.2 and X, and betwee Y.1 and Y using the relational cue "includes", so that X and Y functioned as the most inclusive levels of two separate hierarchical networks. The stimuli were then established as possessing specific functions or characteristics (i.e., X.1 was established as cold, D2 as heavy, and C3 as sweet). Participants were then assessed for the transformation of stimulus function using seven stimuli from both hierarchical networks established within the experiment. Of the six participants who reached this phase of the experiment, five responded correctly.

Both Gil et al. (2012) and Gil et al. (2014) modelled hierarchical responding as broadly conceptualised. As described in Chapter 1, however, mainstream research has suggested a number of varieties of hierarchical responding including, for example, hierarchical classification and hierarchical analysis, and these varieties may have functionally important differences. Accordingly, a number of more recent RFT studies have focused on using an RFT approach to investigate particular varieties of conceptual hierarchy.

Slattery and Stewart (2014) used relational framing to model hierarchical classification specifically and to probe for the three core features said to characterise it, namely, transitive class containment, asymmetrical class containment and unilateral property induction. Phase 1 of their study established arbitrary shapes as contextual cues for hierarchical relational responding, similar to Gil et al. (2012, 2014). In contrast to the latter however, the core of this training involved a very simple set of colored shape stimuli which might be grouped together only along particular physical dimensions. This aimed to facilitate tight control over the non-arbitrary relational pattern supporting the establishing of the contextual cues, so as to facilitate hierarchical classification in particular; more specifically, by inducing relating of abstracted common physical properties (‘classes’) with examples of shapes that included those particular properties (‘members’). In Phase 2, the cues thus established were used to train and test a hierarchical relational network of nonsense syllables. Results were that nine out of ten participants who completed the protocol exhibited predicted patterns of derived relational responding and transformation of function in accordance with all three of the properties of hierarchical classification by showing asymmetrical mutual entailment, transitive combinatorial entailment and unidirectional transformation of function.

The RFT work just reviewed and in particular the study by Slattery and Stewart (2014) is useful in indicating how relational framing may underlie responding in accordance with hierarchical classification. However, this work over the last few years has been conducted with solely adult populations. It remains to be seen whether children show similar patterns of framing, at what age such patterns begin to emerge and whether they can be trained when absent. The research in the current programme aimed to address this gap in the literature by focusing on the early development and training of framing underlying hierarchical classification.

Overview of Studies in the Current Research Programme

The primary research goal of this doctoral programme was to investigate hierarchical classification in young children, both typically developing as well as with developmental delay. Towards this end, the programme began with the development and pilot testing of a protocol for the assessment and training of containment and hierarchical relational responding in young children (Study 1 as reported in Chapter 3). In Studies 2 and 3 (Chapter 4), a protocol was used to train arbitrary containment relational responding in typically developing 5-year-old children. In Studies 4 and 5 (Chapter 5) a protocol was used to train arbitrary hierarchical relational framing in typically developing 6-year-olds. Finally, in Studies 6-8 (Chapter 6) class inclusion was trained in both typically developing children as well as individuals with autism.

Chapter 3 – Study 1. Previous research, primarily within the cognitive and developmental psychology domain, has assessed the development of classification repertoires among young children. The aims of Study 1 were (a) to measure patterns of relational framing linked with categorisation in young typically developing children and (b) to correlate framing performance with linguistic and cognitive performance as measured by standardised instruments.

Chapter 4 – Studies 2 and 3. The aim of Study 2 was to pilot-test a training procedure, based on RFT principles, to assess and teach arbitrary containment relational responding to a typically developing 5-year-old in the context of a multiple baseline across components design. This study was used to determine the feasibility and efficacy of training such repertoires and determine generalisation and maintenance effects of training. Study 3 extended this research by assessing the efficacy of this procedure with three typically developing 5-year-olds in the context of a combined multiple baseline design (across relational components, and across participants). Participants receiving training were also compared to controls at both pre-training and 6 months post-training on a number of pertinent measures of intellectual performance, in order to determine the possible impact of training on the latter.

Chapter 5 – Studies 4 and 5. These studies aimed to access training procedures to facilitate arbitrary hierarchical relational responding in young children. Study 4 pilot-tested a procedure, based on RFT principles, to assess and train arbitrary hierarchical relational responding in a typically developing 6-year-old in the context of a multiple baseline across components design. This study also aimed to assess both generalisation and maintenance of arbitrary hierarchical relational responding. Study 5 expanded on it's predecessor by applying the same arbitrary hierarchy training procedure with a sample of typically developing 6-year-olds in the context of a combined multiple baseline design (across relational components, and across participants). Experimental participants were compared with controls at pre-training and 6 months post-training on several relevant measures to determine the possible impact of the intervention on intellectual performance.

Chapter 6 – Studies 6 - 8. Study 6 used RFT-based procedures to assess and train Piagetian class inclusion responding in typically developing children (aged 3-4) using a multiple baseline across participants design. Study 7 examined the efficacy of this training procedure to facilitate class inclusion responding with autistic individuals in a multiple baseline across participants design. Study 8 extended on the previous two studies by improving the experimental design, introducing participants with multiple and dual diagnoses, modifying the training procedure and expanding on the generalisation assessment procedure employed.

Summary and Conclusion

The current chapter has introduced the RFT approach to hierarchical classification. It started by describing initial empirical work on derived relational responding, before subsequently introducing the functional contextual account of language and cognition offered by RFT, which sees arbitrarily applicable relational responding (AARR) as underlying complex human performance including both derived relational responding and human language. Examples of RFT research teaching various types of relational framing as a means of boosting intellectual performance were provided to indicate the progress of this approach with respect to its central thesis as well as to show its potential for assessing and training both typically developing children and children with developmental disabilities. The chapter then focused on the RFT approach to hierarchical classification, which is the focus of this doctoral research. Finally, a brief overview of the empirical work in the ensuing chapters of the thesis was provided.

In the next chapter the first of the eight studies, on assessing relational framing of hierarchical classification, will be reported.

Chapter 3

Study 1

Investigating Relational Framing of Categorisation in Young Children*

* This chapter has been published:

Mulhern, T., Stewart, I., & McElwee, J. (2017). Investigating relational framing of categorization in young children. *The Psychological Record*. DOI: 10.1007/s40732-017-0255-y

As outlined previously in Chapter 1, the bulk of research examining classification as a skillset has been conducted within the field of cognitive developmental psychology (e.g., Arterberry & Bornstein, 2012; Bornstein & Arterberry, 2010; Blewitt, 1994; Deneault & Ricard, 2006; Gelman, 1988; Greene, 1994; Inhelder & Piaget, 1964; Rostad, Yott, & Poulin-Dubois, 2012; Valentin & Chanquoy, 2012). Research within this field suggests that this skillset is acquired gradually, beginning within early childhood (3 – 4 months; see Arterberry & Kellman, 2016; Eimas & Quinn, 1994; Eimas, Quinn & Cown 1994; Quinn, 2016; Quinn & Eimas, 1996; Song, Pruden, Michnick Golinkoff, Hirsch-Pasek, 2016) with basic classification skills, and culminating later in childhood (approximately 11-years-old; see Blewitt, 1989, 1994; Carneiro, Albuquerque, & Fernandez, 2009; Deneault & Ricard, 2006) with advanced 'hierarchical' classification responding.

Some studies have examined levels of abstraction of stimuli in hierarchical classes, including basic, superordinate and subordinate. For example, 'dog' is at the basic level, 'animal' is at the superordinate level and 'labrador' is at the subordinate level (Mervis & Crisafi, 1982). Evidence suggests that pre-schoolers (i.e., ages 2 to 5), and even potentially infants (i.e., 12 months to 2 years), may be capable of forming certain classes, such as the basic level categories (Blewitt, 1983; Callanan, 1985; Mervis, Johnson, & Mervis, 1994; Murphy, 2016; Smith, Smallman, & Rucker, 2016; Westemann & Mareschal, 2014). There is some evidence to suggest that the acquisition of basic level categorisation occurs prior to the acquisition of superordinate categorisation (Deneault & Ricard, 2005; Horton & Markman, 1980; Shylaja & Manjula, 2016; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 2004). Other studies reveal an alternative developmental order in which children learn superordinate categorisation before they learn categorisation at the basic level (Mandler, Bauer & McDonough, 1991; Mandler & McDonough, 1993; Murphy, 2016; Poulin-Dubois, Graham, & Sippola, 1995; Westemann & Mareschal, 2014). Meantime subordinate categories may provide a greater challenge than either basic level (Hajibayova, 2016; Mervis et al., 1994) or superordinate categories (Blewitt, 1983; Callanan, 1985; Shylaja & Manjula, 2016) for younger children.

Other studies have focused on the properties of hierarchical classification, including asymmetrical and transitive class containment and unilateral property induction (Blewitt, 1994; Deneault & Ricard, 2005; 2006; Diesendruck & Shatz, 2001; Inhelder & Piaget, 1964; Vygotsky, 1962). However, the precise age at which such repertoires emerge have not been pinpointed exactly. Transitive class containment involves inferring that if "X is a member of class Y" and "Y is a member of class Z", then "X is a member of class Z". For example, if labradors are dogs and dogs are animals then labradors are animals. Asymmetrical class containment refers to the derivation that if class Y contains class Z, then class Z does not contain class Y (e.g., if "animal" contains "dog", then "dog" does not contain "animal"). Unilateral property inheritance refers to the inference that the properties of a superordinate class are common to all members of that class (e.g., all animals breathe and therefore all dogs breathe); however, all properties of a subordinate class need not be common to all members of the overarching class (e.g., all labradors are tall but not all dogs are tall).

It has been found that children do not typically show accurate inference regarding hierarchically structured categories until the age of five (Borst, Poirel, Pineau, Cassotti, & Houdé, 2013; Inhelder & Piaget, 1964; Pasnak, Cooke &

Hendricks, 2006; Siegler & Svteina, 2006; Winer, 1980). At this age children begin to show transitive class containment, however, they do not begin to show asymmetry until the age of nine (Deneault & Ricard, 2006; Winer, 1980). These results remain true even when researchers have taken measures to eliminate test biases or have developed alternative measurements (Hodkin, 1987; Markman & Seibert, 1976; McGarrigle, Grieve, & Hughes, 1978; Siegel, McCabe, Brand, & Matthews, 1978; Thomas & Horton, 1997).

Hierarchical classification in children has also been compared with other forms of hierarchical responding. For example, research has compared member/class (class-concept) hierarchical relations with part/whole (collection-concept) hierarchical relations. Class-concept relations rely on common characteristics (e.g., soldiers share the feature of uniforms) whereas collection-concept relations are founded on proximity (e.g., an army is defined by the proximity of numerous soldiers to one another). Markman and Seibert (1976) compared young children's execution of class-concept and collection-concept tests similar to class inclusion. They discovered that younger children exhibited responding exclusively consistent with collection-concept rather than class-concept. The researchers maintain that this indicates that the class-concept is acquired after the development of collection-concept.

While cognitive-developmental research on classification has identified several crucial features of classification and has generated data regarding variables that may affect it, this approach is largely descriptive. The principal aim of such research is to gain a theoretical comprehension of the phenomenon in question, rather than endeavouring to isolate environment-behaviour interactions that can facilitate prediction and influence over behaviour. As such, theoretical accounts offered within the cognitive-developmental approach can be argued to be less than optimal for realising practical change within the applied domain (Margolis & Laurence, 2000; Murphy, 2002; Palmer, 2002).

Behavioural Research and Relational Frame Theory

The approach to classification taken in the current thesis is based on Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001; Dymond & Roche, 2013), which was previously outlined in Chapter 2. This approach conceptualises language and cognition as patterns of generalised contextually controlled relational responding that are learned via multiple exposures to behaviour-environment contingencies operating in the socioverbal environment.

While previous research in the area of RFT has examined relational framing in accordance with hierarchy (e.g., Gil, Luciano, Ruiz & Valdivia-Salas, 2012, 2014; Gil-Luciano, Ruiz, Valdivia-Salas & Suárez-Falcón, 2017), constituting an important starting point with respect to the RFT induction and investigation of hierarchical relational framing, this research has solely been conducted with adult populations. However, such research provides evidence that humans engage in this particular pattern of contextually controlled relational responding and that it can be brought under contextual control in the laboratory. A next step in the empirical investigation of this pattern of relational framing is to begin to examine its historical origins. That was the aim of the current study which sought to measure the emergence of relational framing related to classification, including both containment and hierarchy, in young children spanning the age range of 3 – 8 years and to correlate performance in this regard with performance on standardised measures of intellectual potential.

The current study involved using a protocol developed specifically to track repertoires of relational framing related to hierarchical classification from simple and concrete up to complex and abstract; specifically, non-arbitrary containment, arbitrary containment and arbitrary hierarchy, each assessed in terms of both mutual and combinatorial entailment and, in the case of arbitrary hierarchy, also transformation of function. We anticipated that doing so might provide information regarding typical acquisition of relational framing repertoires relevant to classification. This might further inform researchers and practitioners regarding the development of relational framing in general as well as regarding framing related to hierarchical classification specifically. It might provide information regarding typical hierarchically-relevant framing repertoires at different ages as well as potential deficits at particular ages that might be amenable to training. As indicated earlier, results of previous research suggest that relational framing is positively correlated with performance on measures of intellectual potential. It might be hypothesised that the framing repertoires of containment and hierarchy specifically would be similarly correlated. Assuming that this is the case, such a finding might point to the potential importance of training relational framing skills of categorisation using a protocol based on that developed for the current study.

To assess non-arbitrary containment, we presented children with differently coloured boxes in which smaller boxes were physically contained inside larger boxes and asked the children whether one particular box was inside another or contained the other. To assess arbitrary containment, we presented the children with a number of differently coloured circles that were physically the same size as each other. We then told them about one or more ‘containment’ relations between the circles and probed to see whether they could correctly derive further relations in the absence of any physical containment relations being demonstrated (hence, these tasks probed for arbitrary relations). For example, for mutual entailment tasks, we told them that one particular coloured circle (e.g., the green one) contained another particular coloured circle (e.g., the red one) and then probed to see whether they could derive the correct entailment relation (i.e., in this case, that the red one was inside the green one). Finally, to assess arbitrary hierarchy we presented children with real and nonsense words on a computer screen and told them about one or more ‘hierarchical class’ type relations between the stimuli and probed to see whether they could correctly derive further relations. For example, for mutual entailment tasks, we told them that one particular nonsense word (e.g., ‘tol’) was a type of animal and then probed to see whether they could endorse the correct entailment relation (i.e., in this case, that the class ‘animal’ contained ‘tols’ as members). As for the previous task, this was an arbitrary relational task because the stimuli representing classes and members did not show any physical relationship of relevance to the task.

To assess whether the relational repertoires measured in this study correlated with intellectual performance, a number of standardised tests of intellectual skill were employed including the Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn, 2007), the Stanford Binet 5th Edition (SB5; Roid, 2003) and the Children’s Category Test (CCT; Boll, 1993). A test of Piagetian class inclusion (Piaget, 1952) was also employed, to examine whether a particular scoring pattern on the relational framing protocol might predict performance on this well-known test of classification.

Study 1

Method

Participants

Fifty typically developing children (23 female, 27 male) between 3 and 8 years of age (range of 3 years and 0 months to 7 years and 9 months) took part. The mean age and standard deviation for each group (n = 10 each) were as follows: 3-year-olds (*M* = 3.56 years; *SD* = 0.26 years); 4-year-olds (*M* = 4.56 years; *SD* = 0.27 years); 5-year-olds (*M* = 5.54 years; *SD* = 0.29 years); 6-year-olds (*M* = 6.31 years; *SD* = 0.28 years); and 7-year-olds (*M* = 7.27 years; *SD* = 0.27 years).

Participants were recruited from rural primary schools and play schools within the West of Ireland. Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland, Galway, Research Ethics Committee. Consent for conducting the study was obtained from the principal in each respective school. Parental consent was obtained for each child who participated and verbal consent was also obtained from each of the participants.

Materials

Participants were exposed to a number of different assessments. These included pre-assessment of (a) colour tacting and (b) yes/no responding. They were also assessed on a number of standardised measures as follows: (a) the Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn, 2007); (b) the Stanford Binet 5 (SB5; Roid, 2003); and (c) the Children's Category Test (CCT; Boll, 1993). The assessments also included a test of class inclusion (Piaget, 1952) and a test of relational responding involving three parts: (a) non-arbitrary containment; (b) arbitrary containment; and (c) arbitrary hierarchy.

Colour tacting. This test involved nine paper circles each of which was coloured with a different colour used during the relational assessment procedure (i.e., black, white, orange, green, blue, yellow, purple, red, or pink). During the test, the child was exposed to three arrays of three coloured circles each and asked to tact each particular colour displayed. Children had to answer correctly on all nine trials to proceed.

Yes - No responding. This test involved 10 pictures of commonly seen objects or animals (e.g., a dog). On each trial the child was shown a picture of an object or animal and asked a yes-no question. Children were asked five questions to which the answer *yes* was appropriate (e.g., they were shown a picture of a dog and asked “Is this a dog?”) and five questions to which the answer *no* was appropriate (e.g., they were shown a picture of a circle and asked “Is this a square?”). Children had to answer correctly on all 10 trials to proceed.

PPVT-4. This test assesses receptive vocabulary in respondents from ages 2 years and 6 months to late adulthood and is often used as a test of scholastic aptitude. It is administered individually and is presented in a multiple-choice format in which the respondent is presented with four pictures, and asked to select the one that best illustrates the definition of a particular word. The stimuli include items representing up to 20 content areas (e.g., actions, vegetables, tools) and components of speech (nouns, verbs, or attributes), encompassing a broad range of difficulty.

Test-retest reliability has been shown to be between .92 and .96 (Community-University Partnership for the Study of Children, Youth, and

Families, 2011), and internal consistency has been noted to be similarly high (between .94 and .95). Construct and convergent validity has been assessed by comparing the PPVT-4 to the Expressive Vocabulary Test, Second Edition (EVT-2; Williams, 2007). Correlations between the two were high (r = .80 to .84) for all age groups (Community-University Partnership for the Study of Children, Youth, and Families, 2011).

SB5. This is an intelligence test for use with individuals from ages 2 to 85 years that measures five weighted factors (fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing and working memory) and consists of both verbal and non-verbal subtests. Participants in the current study were evaluated for the full-scale intelligence quotient (FSIQ), the verbal intelligence quotient (VIQ), and the non-verbal intelligence quotient (NVIQ). Reliability coefficients have been found to be extremely high for the FSIQ (r = .98), the NVIQ (r = .95), and the VIQ (r = .96), showing excellent stability. In addition, the five factor index scores were all above .90, and were higher than the subtest scales, which were however comparable to other cognitive tests with ranges from .84 to .89 (Roid, 2003). Validity has been assessed through comparison with other tests of cognitive and intellectual ability, including the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989), the Wechsler Intelligence Scale for Children, third edition (WISC-III; Wechsler, 1991), the Wechsler Adult Intelligence Scale, third edition (WAIS-III; Wechsler, 1997), the Woodcock-Johnson III Test of Cognitive Ability, and the Woodcock-Johnson III Test of Achievement. Correlations ranged from .78 to .84 for both FSIQ and VIQ with comparable indices on other major IQ batteries (Roid, 2003).

CCT. This test was designed to assess categorisation skills for children between 5 and 8 years of age (Level 1) and between 9 and 16 years of age (Level 2). It is individually administered and is designed to assess concept formation and problem-solving abilities with novel material. In the current study, only Level 1 was used. This level consists of five subtests and includes 80 questions. Within each subtest children are asked to determine the principle underlying correct performance (i.e., matching based on colour, size or proportion) using examiner feedback. Previous research has indicated that the test has adequate test-retest reliability (r = .75), in addition to strong internal consistency for both Levels 1 (r = .88) and 2 (r = .86; MacNeill, 1996).

Piagetian class inclusion. This is a classic test of categorisation ability that assesses the capacity to respond to a stimulus as simultaneously belonging to both a class and a superordinate class (Thomas & Horton, 1997). It has its origins in Piagetian developmental psychology (Piaget, 1952) where it was seen as determining if a child had reached the "concrete operational" stage of development, hypothesised to occur by age 7 or 8. In the test, the child is first shown an array of stimuli that belong to a particular category. These stimuli form two different subclasses, with a greater number of stimuli within one category than the other (e.g., five apples and two pears). The child is asked if there are more of the larger subclass or of the category; for example – "Are there more apples or is there more fruit?" Materials used in the current study included 5.5cm x 5.5cm coloured flashcards depicting a variety of animals (horse, dog, pig, cow, cat, and dog), fruit (strawberry, apple, pear, banana, lemon, orange), clothing (socks, skirt, t-shirt, dress, pants, coat) and vehicles (car, fire engine, truck, motorcycle, tractor, bus). Multiple examples of each of these stimuli were used.

Relational responding test. The current study assessed three different patterns of relational responding including (a) non-arbitrary containment, (b) arbitrary containment, and (c) arbitrary hierarchy. They were assessed as follows:

1. Non-arbitrary containment: Materials included both square and triangular boxes of different sizes (large, medium, or small) and colours (red, yellow, green, blue, purple, orange, black, white, and pink).

2. Arbitrary containment: Materials included same-sized circles and same-sized triangles of differing colours (red, yellow, green, blue, purple, orange, black, white, and pink).

3. Arbitrary hierarchy: This test used a laptop to show arrays of stimuli that were used as the basis for asking particular questions to probe for derived relations.

Procedure

In the case of each school setting, the research was conducted in a separate classroom or resource room within the school building. In the first (screening) session, children were exposed to the colour tacting and yes-no pre-tests, which together took a total of 10 min per child. All the children passed these tests. Thereafter, they were tested individually on the main assessments in sessions lasting between 30 and 45 min, with the length of a session depending on the test being administered and the age of the child. These sessions involved breaks every 10 or 15 min, also depending on the age of the child. Children were exposed to the main assessments in the following order, with the number of sessions involved for each measure and the range of testing time per session, including breaks, given in parentheses:

1. Relational Responding Test 1 (3 sessions lasting 40-45 minutes each, one for non-arbitrary containment, one for arbitrary containment and one for arbitrary hierarchy).

2. SB5 (two to three sessions, with 40 min for Session 1, 25-30 min for Session 2 and 20-25 min for Session 3, with the number of sessions and exact duration depending on the age of the child).

3. PPVT-4 (one session, 30-40 min).

4. CCT (one session, 30 min).

5. Class inclusion (included in the same session as the CCT, 10 min).

6. Relational Responding Test 2 (number and length of sessions were the same as for Relational Responding Test 1).

Testing sessions took place over two phases. Phase 1, which included Assessments 1 – 5, involved approximately seven to eight days of testing per child spread out over a duration of approximately 2 weeks. Phase 2, which included Assessment 6 (i.e., the second relational responding test) took place 2 weeks after the initial phase ended and involved a successive 3 days of testing.

Relational responding test. In their first session after passing the screening tests, participants were assessed for the three relational responding repertoires seen as involved in classification, namely, non-arbitrary containment (i.e., relating stimuli on the basis of an observed physical containment relationship), arbitrary containment (i.e., relating stimuli on the basis of an abstract or arbitrary containment relationship), and arbitrary hierarchy (i.e., relating stimuli on the basis of an abstract hierarchical class relationship). These three relational repertoires were assessed over the course of three sessions (one repertoire per session) based on 128 questions (32 for each of the first two repertoires and 64 for the third).

In the case of non-arbitrary and arbitrary containment, relational framing was assessed in terms of mutual (16 questions) and combinatorial entailment (16 questions (see Tables 3.1 and 3.2 respectively). On all trials, participants were first shown the relevant stimuli, and then the experimenter described the relationship between the stimuli. For the assessment of non-arbitrary containment responding, the experimenter further demonstrated the relationship by manipulating the stimuli (e.g., following the description “A red box is inside a blue box”, the experimenter would then place a red box inside a blue box). Participants were then asked a question about the relationship between certain stimuli that could typically be answered in the form of a *yes* or *no* response (more specific details regarding format are provided later in subsections corresponding to each repertoire). No contingent feedback was provided at any time.

In the case of arbitrary hierarchy, relational framing was assessed in terms of mutual entailment (16 questions), transformation of functions through mutual entailment (16 questions), combinatorial entailment (16 questions) and transformation of functions through combinatorial entailment (16 questions). On all trials, participants were first shown an array of stimuli on a computer screen, and then the experimenter described the relationship between the stimuli, elaborating slightly for transformation of function questions. As in previous sections, participants were then presented with a follow-up question that could typically be answered in the form of a “yes” or “no” response (again, more specific details regarding format are provided in later section) and no contingent feedback was provided at any time (see Appendix A for arbitrary hierarchy assessments).

As described previously, for all three relational repertoires, questions were scored in groups of 16. In the case of each item in each 16-question section, participants received either 1 (correct) or 0 (incorrect) for that item and thus they received a score from 0 to 16 for that section. A score of 13 out of 16 or higher was deemed a pass for that section (it was calculated that achieving a score within this range by chance was approximately 1 in 100). Section final scores were added to give cumulative scores for each relational repertoire and an overall relational framing score.

Participants were first assessed for non-arbitrary containment responding across 32 trials. Table 3.1 shows a generic representation of the trial types involved. In tests for mutual entailment, the participant was presented with two differently coloured boxes (one inside another), and all questions within this phase of assessment were focused on the relationship between these two stimuli. A total of 16 questions (including eight questions that could be answered *yes* and eight that could be answered *no*) were presented in random order during this phase. In tests for combinatorial entailment, the participant was presented with three differently coloured boxes (one inside a second and the second inside a third), and all 16 questions in this phase focused on the relationship between the first and third boxes (See Table 3.1).

Table 3.1. Trial types used in the non-arbitrary containment phase.

Property	Statement	Question [Answer]
Mutual Entailment (2 stimuli)	The [A] box is inside the [B] box	Is the [A] box inside the [B] box? [Yes]
	The [A] box is inside the [B] box	Is the [B] box inside the [A] box? [No]
	The [A] box is inside the [B] box	Does the [A] box contain the [B] box? [No]
	The [A] box is inside the [B] box	Does the [B] box contain the [A] box? [Yes]
	The [B] box is inside the [A] box	Is the [A] box inside the [B] box? [No]
	The [B] box is inside the [A] box	Is the [B] box inside the [A] box? [Yes]
	The [B] box is inside the [A] box	Does the [A] box contain the [B] box? [Yes]
	The [B] box is inside the [A] box	Does the [B] box contain the [A] box? [No]
	The [A] box contains the [B] box	Is the [A] box inside the [B] box? [No]
	The [A] box contains the [B] box	Is the [B] box inside the [A] box? [Yes]
	The [A] box contains the [B] box	Does the [A] box contain the [B] box? [Yes]
	The [A] box contains the [B] box	Does the [B] box contain the [A] box? [No]
	The [B] box contains the [A] box	Is the [A] box inside the [B] box? [Yes]
	The [B] box contains the [A] box	Is the [B] box inside the [A] box? [No]
	The [B] box contains the [A] box.	Does the [A] box contain the [B] box? [No]
	The [B] box contains the [A] box	Does the [B] box contain the [A] box? [Yes]
Comb. Entailment (3 stimuli)	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Is the [A] box inside the [C] box? [Yes]
	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Is the [C] box inside the [A] box? [No]
	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Does the [A] box contain the [C] box? [No]
	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Does the [C] box contain the [A] box? [Yes]
	The [C] box is inside the [B] box; The [B] box is inside the [A] box	Is the [A] box inside the [C] box? [No]
	The [C] box is inside the [B] box The [B] box is inside the [A] box	Is the [C] box inside the [A] box? [Yes]
	The [C] box is inside the [B] box; The [B] box is inside the [A] box	Does the [A] box contain the [C] box? [Yes]
	The [C] box is inside the [B] box; The [B] box is inside the [A] box	Does the [C] box contain the [A] box? [No]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Is the [A] box inside the [C] box? [No]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Is the [C] box inside the [A] box? [Yes]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Does the [A] box contain the [C] box? [Yes]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Does the [C] box contain the [A] box? [No]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Is the [A] box inside the [C] box? [Yes]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Is the [C] box inside the [A] box? [No]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Does the [A] box contain the [C] box? [No]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Does the [C] box contain the [A] box? [Yes]

Note. The codes [A], [B] and [C] signify different colours within each trial. The actual colours of the box stimuli (e.g., red, green, blue, yellow, pink, purple etc.) varied quasi-randomly across trials.

Participants were next assessed for arbitrary containment responding across 32 trials. The trial types for this assessment were similar to those for the assessment of non-arbitrary containment, but this assessment used identically sized coloured-circle stimuli between which no physical containment relationship was demonstrated (see Table 3.2 for a generic representation of the trial types involved).In tests for mutual entailment of arbitrary containment, the participant was presented with two shapes (e.g., circles) of equal size, but different colour, and all questions in this assessment phase were focused on the relationship between these two stimuli. A total of 16 questions, presented in random order, were presented during this phase. In tests for combinatorial entailment, the participant was presented with three shapes (designated as A, B, and C) of equal size, but different colour, and all 16 questions in this phase focused on the derivation of a relation between Stimuli A and C based on the combination of given relations between A and B and between B and C, respectively.

In the third session of relational testing, participants were assessed for arbitrary hierarchical relational responding across a total of 64 questions (see Table 3.3 for a list of the trial types involved).

In all tests for arbitrary hierarchy, the participant was presented (on a laptop computer) with on-screen textual descriptions of the relationship(s) among one or more pairs of stimuli (see Figure 3.1), and these descriptions were also read aloud. The questions in this phase were presented in groups of eight items each (see Table 3.3); the questions involved were presented in random order in each. The groups were presented in the following order:

1. Mutual entailment (ME) questions including the cue *type of*.

2. ME questions including the cue *type of* that assessed transformation of functions (ToF).

3. ME questions including the cue *contains*.

4. ME questions including the cue *contains* that assessed ToF.

5. Combinatorial entailment (CE) questions including the cue *type of*.

6. CE questions including the cue *type of* that assessed ToF.

Table 3.2. Trial types used in the arbitrary containment phase.

Property	Statement	Question [Answer]
Mutual Entailment (2 stimuli)	The [A] circle is inside the [B] circle	Is the [A] circle inside the [B] circle? [Yes]
	The [A] circle is inside the [B] circle	Is the [B] circle inside the [A] circle? [No]
	The [A] circle is inside the [B] circle	Does the [A] circle contain the [B] circle? [No]
	The [A] circle is inside the [B] circle	Does the [B] circle contain the [A] circle? [Yes]
	The [B] circle is inside the [A] circle	Is the [A] circle inside the [B] circle? [No]
	The [B] circle is inside the [A] circle	Is the [B] circle inside the [A] circle? [Yes]
	The [B] circle is inside the [A] circle	Does the [A] circle contain the [B] circle? [Yes]
	The [B] circle is inside the [A] circle	Does the [B] circle contain the [A] circle? [No]
	The [A] circle contains the [B] circle	Is the [A] circle inside the [B] circle? [No]
	The [A] circle contains the [B] circle	Is the [B] circle inside the [A] circle? [Yes]
	The [A] circle contains the [B] circle	Does the [A] circle contain the [B] circle? [Yes]
	The [A] circle contains the [B] circle	Does the [B] circle contain the [A] circle? [No]
	The [B] circle contains the [A] circle	Is the [A] circle inside the [B] circle? [Yes]
	The [B] circle contains the [A] circle	Is the [B] circle inside the [A] circle? [No]
	The [B] circle contains the [A] circle.	Does the [A] circle contain the [B] circle? [No]
	The [B] circle contains the [A] circle	Does the [B] circle contain the [A] circle? [Yes]
Comb. Entailment (3 stimuli)	The [A] circle is inside the [B] circle; the [B] circle is inside the [C] circle	Is the [A] circle inside the [C] circle? [Yes]
	The [A] circle is inside the [B] circle; The [B] circle is inside the [C] circle	Is the [C] circle inside the [A] circle? [No]
	The [A] circle is inside the [B] circle; The [B] circle is inside the [C] circle	Does the [A] circle contain the [C] circle? [No]
	The [A] circle is inside the [B] circle; The [B] circle is inside the [C] circle	Does the [C] circle contain the [A] circle? [Yes]
	The [C] circle is inside the [B] circle; The [B] circle is inside the [A] circle	Is the [A] circle inside the [C] circle? [No]
	The [C] circle is inside the [B] circle The [B] circle is inside the [A] circle	Is the [C] circle inside the [A] circle? [Yes]
	The [C] circle is inside the [B] circle; The [B] circle is inside the [A] circle	Does the [A] circle contain the [C] circle? [Yes]
	The [C] circle is inside the [B] circle; The [B] circle is inside the [A] circle	Does the [C] circle contain the [A] circle? [No]
	The [A] circle contains the [B] circle; The [B] circle contains the [C] circle	Is the [A] circle inside the [C] circle? [No]
	The [A] circle contains the [B] circle;	Is the [C] circle inside the [A] circle? [Yes]

	The [B] circle contains the [C] circle	
	The [A] circle contains the [B] circle; The [B] circle contains the [C] circle	Does the [A] circle contain the [C] circle? [Yes]
	The [A] circle contains the [B] circle; The [B] circle contains the [C] circle	Does the [C] circle contain the [A] circle? [No]
	The [C] circle contains the [B] circle; The [B] circle contains the [A] circle	Is the [A] circle inside the [C] circle? [Yes]
	The [C] circle contains the [B] circle; The [B] circle contains the [A] circle	Is the [C] circle inside the [A] circle? [No]
	The [C] circle contains the [B] circle; The [B] circle contains the [A] circle	Does the [A] circle contain the [C] circle? [No]
	The [C] circle contains the [B] circle; The [B] circle contains the [A] circle	Does the [C] circle contain the [A] circle? [Yes]

Note. The codes [A], [B] and [C] signify different colours within each trial. The actual colours of the circle stimuli (e.g., red, green, blue, yellow, pink, purple etc.) varied quasi-randomly across trials.

7. CE questions including the cue *contains*.

8. CE questions including the cue *contains* that assessed ToF.

In each item in the basic ME phases (i.e., Groups 1 and 3), an initial statement involving a hierarchical relational cue was presented (e.g., “A tol is a type of animal”), and the participant was then required to respond correctly to a follow-up question based on this provided statement (see Table 3.3). Correctly answered items in Groups 1 and 3 were combined to give a score for ME. In each item in the ME ToF phases (i.e., Groups 2 and 4), an initial statement involving a hierarchical relational cue was presented, a function (e.g., “big eyes”) was (verbally) associated with one or other of the two stimuli, and the participant was required to respond correctly to a follow-up question based on this. Correctly answered items in Groups 2 and 4 were combined to give a score for ME with ToF.

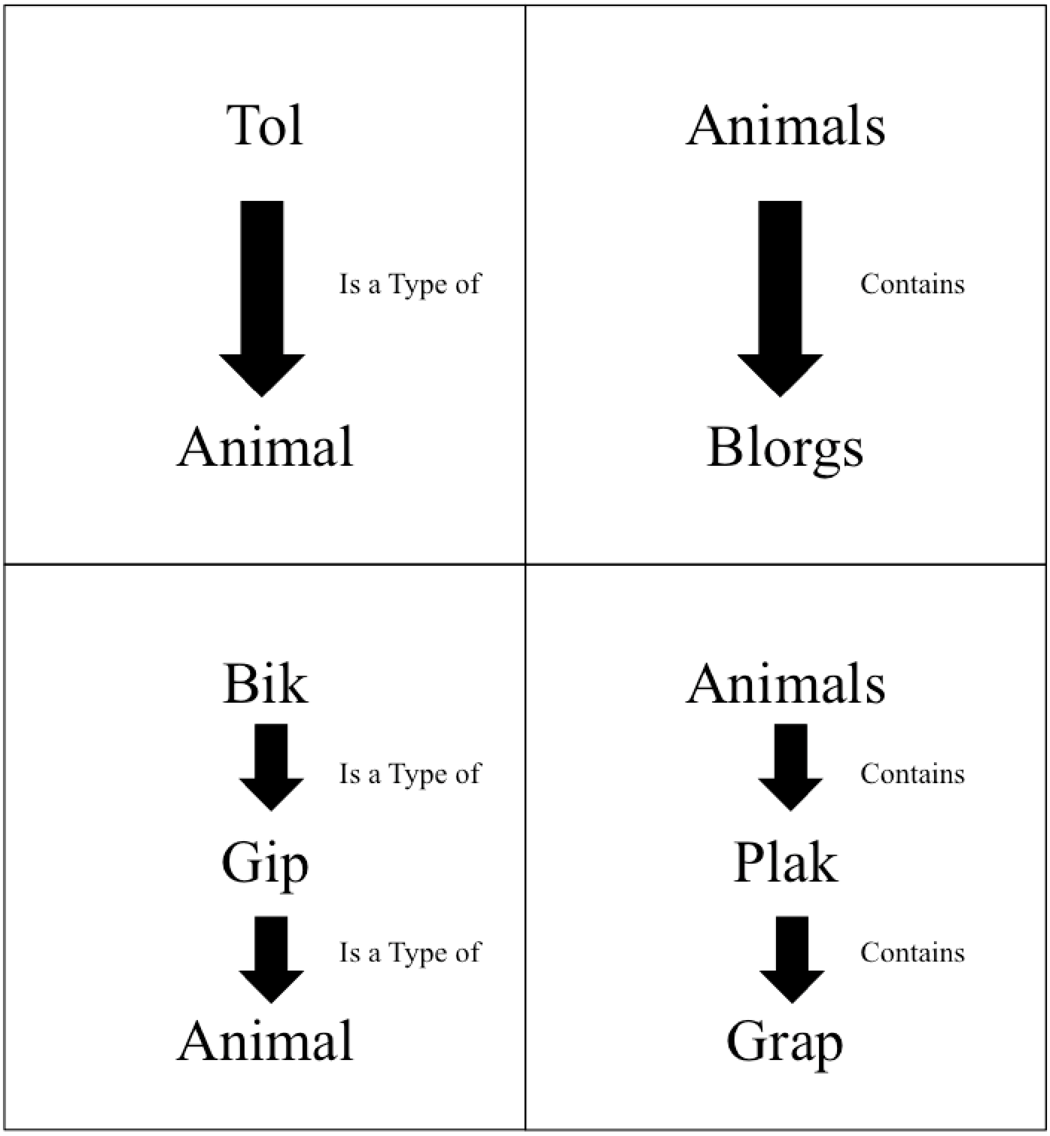
Table 3.3. Trial types used in the arbitrary hierarchy phase.

Property	Statement	Question [Answer]
ME (2 stimuli)	A tol is a type of animal	Is a tol a type of animal? [Yes] Is an animal a type of tol? [No]
	A blorg is a type of animal.	Are all blorgs animals? [Yes] Are all animals blorgs? [No]
	A zimp is a type of animal.	Are some animals zimps? [Yes] Does the class of zimps contain all of the animals? [No]
	A quig is a type of animal.	Does the class of animals contain quigs? [Yes] Are all animals quigs? [No]
ToF (2 stimuli)	A tol is a type of gip. Gips have bones.	Do all tols have bones? [Yes] Does the class of tols contain members without bones? [No]
	A blorg is a type of grap. Graps have freckles.	Does the class of blorgs contain members that have freckles? [Yes] Does the class of graps contain members without freckles? [No]
	A zimp is a type of crint. Zimps have spots.	Do all crints have spots? [Don't Know] Do some crints have spots? [Yes]
	A quig is a type of donk. Quigs have purple tongues.	Does the class of donks contain members with purple tongues? [Yes] Do all donks have purple tongues? [Don't Know]
ME (2 stimuli)	The class of animals contains tols.	Are all tols animals? [Yes] Are all animals tols? [No]
	The class of animals contains blorgs.	Does the class of animals contain blorgs? [Yes] Does the class of blorgs contain all of the animals? [No]
	The class of animals contains zimps.	Does the class of animals contain zimps? [Yes] Are all animals zimps? [No]
	The class of animals contains quigs.	Are some animals quigs? [Yes] Does the class of quigs contain all of the animals? [No]
ToF (2 stimuli)	The class of crints contains zimps. Crints are furry.	Does the class of zimps contain members that are furry? [Yes] Does the class of crints contain members that aren't furry? [No]
	The class of donks contains quigs. Donks have square eyes.	Do all quigs have square eyes? [Yes] Does the class of quigs contain members without square eyes? [No]
	The class of gips contains tols. Tols have big green eyes.	Does the class of gips contain members with big green eyes? [Yes] Do some gips have big green eyes? [Yes]
	The class of graps contains blorgs. Blorgs have whiskers.	Do all graps have whiskers? [Don't Know] Do some graps have whiskers? [Yes]

CE (3 stimuli)	A bik is a type of gip, a gip is a type of animal.	Is a bik a type of animal? [Yes] Is an animal a type of bik? [No]
	A grap is a type of plak, a plak is a type of animal.	Are all graps animals? [Yes] Are all animals graps? [No]
	A crint is a type of wark, a wark is a type of animal.	Are some animals crints? [Yes] Does the class crint contain all of the animals? [No]
	A donk is a type of frik, a frik is a type of animal.	Does the class of animals contain donks? [Yes] Are all animals donks? [No]
ToF (3 stimuli)	A bik is a type of gip, a gip is a type of timp. Timps like sweet food.	Do all biks like sweet food? [Yes] Does the class of timps contain members who don't like sweet food? [No]
	A rab is a type of grap, a grap is a type of plak. Plaks have scales.	Does the class of rabs contain members that have scales? [Yes] Does the class of rabs contain members without scales? [No]
	A dring is a type of crint, a crint is a type of wark. Drings have tails.	Do all warks have tails? [Don't Know] Do some warks have tails? [Yes]
	A slib is a type of donk, a donk is a type of frik. Slibs have spikes.	Does the class of frik contain members with spikes? [Yes] Do all friks have spikes? [Don't Know]
CE (3 stimuli)	The class of animals contains gips, the class of gips contains biks.	Are all biks animals? [Yes] Are all animals biks? [No]
	The class of animals contains plaks, the class of plaks contains graps.	Is a grap a type of animal? [Yes] Is an animal a type of grap? [No]
	The class of animals contains warks, the class of warks contains crints.	Does the class of animals contain crints? [Yes] Are all animals crints? [No]
	The class of animals contains friks, the class of friks contains donks.	Are some animals donks? [Yes] Does the class donk contain all of the animals? [No]
ToF (3 stimuli)	The class of warks contains crints, the class of crints contains drings. Warks have gold eyes.	Does the class of drings contain members with gold eyes? [Yes] Does the class of warks contain members without gold eyes? [No]
	The class of friks contains donks, the class of donks contains slibs. Friks are cold.	Are all slibs cold? [Yes] Does the class of slibs contain members that aren't cold? [No]
	The class of timps contains gips, the class of gips contains biks. Biks have small noses.	Does the class of timps contain members with small noses? [Yes] Do some timps have small noses? [Yes]
	The class of plaks contains graps, the class of graps contains rabs. Rabs have big ears.	Do all plaks have big ears? [Don't Know] Do some plaks have big ears? [Yes]

Note. ME = Mutual Entailment; ToF = Transformation of Function; CE = Combinatorial Entailment.

Figure 3.1. Sample stimuli layouts for mutual entailment (top panels) and combinatorial entailment (bottom panels) trials in the arbitrary hierarchy testing phases.



In each item in the basic CE phases (i.e., Groups 5 and 7), two statements involving a hierarchical relational cue were presented, and the participant was then required to respond correctly to a follow-up question based on the combination of those statements. Correctly answered items in Groups 5 and 7 were combined to give a score for CE. In each item in the CE ToF phases (i.e., Groups 6 and 8), two statements involving a hierarchical relational cue were presented, a function was associated with one of the stimuli involved, and the participant was required to respond correctly to a follow-up question based on this. Correctly answered items in Groups 6 and 8 were combined to give a score for CE with ToF.

As noted previously in the first paragraph of the procedure, two versions of the relational responding test were given: Relational Responding Tests 1 and 2. Test 2 was similar to Test 1, except that a different set of stimuli was involved. More specifically, in the non-arbitrary and arbitrary containment tests, differently shaped boxes and different shapes, respectively, were used, whereas in the arbitrary hierarchy test, a different set of textual stimuli (i.e., nonsense and real words) was presented (see Appendix A for full set of arbitrary hierarchy re-test questions).

Class inclusion. This test involved presenting eight questions focused on class inclusion (CI) as previously described. These included the following question types:

1. Are there more [smaller subclass] or more [category]?
2. Are there more [larger subclass] or more [category]?
3. Are there more [category] or more [smaller subclass]?
4. Are there more [category] or more [larger subclass]?
5. Are there less [smaller subclass] or less [category]?
6. Are there less [larger subclass] or less [category]?
7. Are there less [category] or less [smaller subclass]?
8. Are there less [category] or less [larger subclass]?

These questions were presented in random order based on the shuffling of eight question cards. Participants received non-contingent reinforcement for taking part.

Inter-Observer Agreement (IOA) and Procedural Fidelity

IOA was conducted by research assistants in the case of 20% of participants and was collected during non-arbitrary containment, arbitrary containment, and arbitrary hierarchy assessment sessions. Prior to data collection, IOA collectors were trained in data collection until they reached 100% accuracy. Mean IOA was calculated as 99.48% (range 96.88% – 100%). Procedural fidelity was conducted for 20% of participants, and was collected during non-arbitrary containment, arbitrary containment, and arbitrary hierarchy assessment sessions. A trained research assistant collected procedural fidelity measures. Prior to data collection, correct experimenter procedural responding was modelled to the data collectors for exemplars of procedural integrity, and data collectors were also provided with a printed copy of the assessment protocol and a checklist of the steps in the assessment procedure. Mean procedural integrity was calculated as 99.56% (range 93.75% - 100%).

Results

Correlations

Table 3.4 shows a correlation matrix of Spearman's ρ correlations among the experimental measures and some of their key subscales.

This table shows a highly significant correlation between age in months and overall relational framing score ($\rho = .852$, $p < .001$), as well as between age and each of the three specific relational repertoires including non-arbitrary containment (ρ = .826, p < .001), arbitrary containment (ρ = .806, p < .001), and arbitrary hierarchy (ρ = .789, p < .001). The data also show highly significant correlations between the overall and specific relational framing scores and those of the other measures and in the case of the SB5, its subscales (i.e., SB5, SB5 Verbal, SB5 Non-verbal, PPVT-4, CCT, and CI). In general the highest correlations are seen for the measures of general intellectual performance (i.e., SB5 and subscales) and language (PPVT-4) with slightly lower correlations for the measures of categorisation (i.e., CCT and CI).

Regarding the correlations among the non-relational framing measures themselves, we can see that the standardised measures of intellectual performance and language correlate very highly with each other, as might be expected (e.g., SB5 and PPVT-4 show a correlation of ρ = .918). Each of these measures also correlates well with the CCT (e.g., PPVT-4: ρ = .769; SB5: ρ = .817), and not quite as well with the CI (e.g., PPVT-4: ρ = .415; SB5: ρ = .432). With regard to the two latter (categorisation) measures: (a) The CCT shows much higher levels of correlation with the measures of intellectual and language performance than the CI; (b) both tests show comparable levels of correlation with the relational framing measures; and (c) the level of correlation between these tests themselves is the lowest in the table (i.e., ρ = .285).

Relational Framing Performance per Age Cohort

Table 3.5 shows the average number of correct responses and corresponding percentage correct responses per age group and per relational framing pattern including both relational frames (non-arbitrary containment, arbitrary containment and arbitrary hierarchy) and relational properties (ME, CE and ToF). The broad patterns seen are in accordance with might have been predicted in that:

1. For each relational property (and thus also all relations) all age cohorts perform in the aggregate at least as well as any younger cohort (with one exception, which is for ME of non-arbitrary containment relations as shown by the 5- to 6-year-old and 6- to 7-year-old groups and in that case the difference is relatively small). In general, aggregate performance on all indices improves with age.

2. For all cohorts and for all three relations, the aggregate score for any CE test is at most as high and is typically lower than the scoring for the corresponding ME test.

Table 3.6 shows the aggregate number of passes achieved within each age cohort for each of the relational properties assessed for each of the three patterns of relational framing (see also Figure 3.2). The broad patterns seen are similar to those seen for the raw data in that:

1. For each relational property (and thus also all relations) all cohorts perform at least as well as any younger cohort and in general performance on all indices improves with age.

2. For all cohorts and all three relations, the score for CE is at most as high and is typically lower than that for ME (with one slightly exception, namely ToF for 6- to 7-year-olds).

Table 3.4. Matrix of Spearman's rho correlations for all measures administered.

	Total RF	**NAC**	**AC**	**AH**	**AM**	**SB5-T**	**SB5-V**	**SB5-NV**	**PPVT**	**CCT**	**CI**
Total RF		.915***	.930***	.960***	.852***	.832***	.778***	.845***	.785***	.615***	.634***
NAC	.915***		.799***	.804***	.826***	.784***	.741***	.790***	.745***	.551***	.544***
AC	.930***	.799***		.870***	.806***	.758***	.705***	.801***	.679***	.605***	.674***
AH	.960***	.804***	.870***		.789***	.770***	.721***	.771***	.731***	.556***	.621***
AM	.852***	.826***	.806***	.789***		.844***	.822***	.850***	.828***	.679***	.605***
SB5-T	.832***	.784***	.758***	.770***	.844***		.966***	.927***	.918***	.817***	.432**
SB5-V	.778***	.747***	.705***	.721***	.822***	.966***		.822***	.891***	.802***	.401**
SB5-NV	.845***	.790***	.801***	.771***	.850***	.927***	.822***		.850***	.756***	.497***
PPVT	.785***	.745***	.679***	.731***	.828***	.918***	.891***	.850***		.769***	.415**
CCT	.615***	.551***	.605***	.556***	.679***	.817***	.802***	.756***	.769***		.285*
CI	.634***	.544***	.674***	.621***	.605***	.432**	.401**	.497***	.415**	.285*	

Note. All scores were raw (unstandardised) scores. Total RF = Total relational framing score; NAC = Total non-arbitrary containment score; AC = Total arbitrary containment score; AH = Total arbitrary hierarchy score; AM = Age in months; SB5-T = Stanford-Binet 5th edition total score; SB5-V = Stanford-Binet 5th edition total verbal score; SB5-NV = Stanford Binet 5th edition total non-verbal score; PPVT = Peabody Picture Vocabulary Test total score; CCT = Children's Category Test total score; CI = Total Class Inclusion score; * = $p < .05$; ** = $p <.01$; *** $p = < .001$

Table 3.5. Average correct response number and percentage per age group and relational framing pattern (including frames and properties).

Age Cohort	Non-Arbitrary Containment			Arbitrary Containment			Arbitrary Hierarchy				
	ME (%)	CE (%)	Total (%)	ME (%)	CE (%)	Total (%)	ME (%)	ToF-M (%)	CE (%)	ToF-C (%)	Total (%)
3–4 years	8 (50%)	7.6 (48%)	15.6 (49%)	6.7 (42%)	5.8 (36%)	12.5 (39%)	6 (37%)	5.9 (37%)	5.6 (35%)	3.8 (24%)	21.3 (33%)
4–5 years	11.2 (70%)	9.2 (58%)	20.4 (64%)	9.8 (61%)	8.3 (52%)	18.1 (57%)	8.3 (52%)	7.1 (44%)	6.7 (42%)	5.2 (33%)	27.2 (43%)
5–6 years	14.2 (89%)	11.6 (73%)	25.8 (81%)	10.4 (65%)	8.4 (53%)	18.8 (59%)	9.4 (59%)	7.6 (47%)	8.4 (53%)	7.5 (47%)	32.8 (51%)
6–7 years	13.6 (85%)	13.3 (83%)	26.9 (84%)	12.8 (80%)	9.8 (61%)	22.6 (71%)	9.8 (61%)	8.7 (54%)	8.5 (53%)	8.1 (51%)	35.1 (55%)
7–8 years	16 (100%)	15.6 (98%)	31.6 (99%)	14.2 (89%)	13 (81%)	27.2 (85%)	13.5 (84%)	11 (69%)	12.3 (77%)	10.9 (68%)	47.7 (75%)

Note. ME = Mutual entailment (no. correct responses out of 16); % = CE = Combinatorial entailment (no. correct responses out of 16); Total = total number of correct entailment responses for a particular frame (no. correct responses out of 32); ToF-M = Transformation of function of mutual entailment (no. correct responses out of 16); ToF-C= Transformation of function of combinatorial entailment (no, correct responses out of 16).

Table 3.6. Number of passes for each relational test per age cohort.

Age Cohort	Non-Arbitrary Containment		Arbitrary Containment		Arbitrary Hierarchy			
	ME	CE	ME	CE	ME	ToF-M	CE	ToF-C
3 – 4yrs	0	0	0	0	0	0	0	0
4 – 5yrs	2	0	0	0	0	0	0	0
5 – 6yrs	7	4	2	0	2	0	0	0
6 – 7yrs	7	7	7	0	2	0	1	1
7 – 8yrs	10	10	10	5	6	2	5	1

Note. ME = Mutual Entailment; ToF-M = Transformation of function for Mutual Entailment; CE = Combinatorial entailment; ToF-C = Transformation of function for Combinatorial Entailment.

Figure 3.2. Number of passes per relational frame and age group.

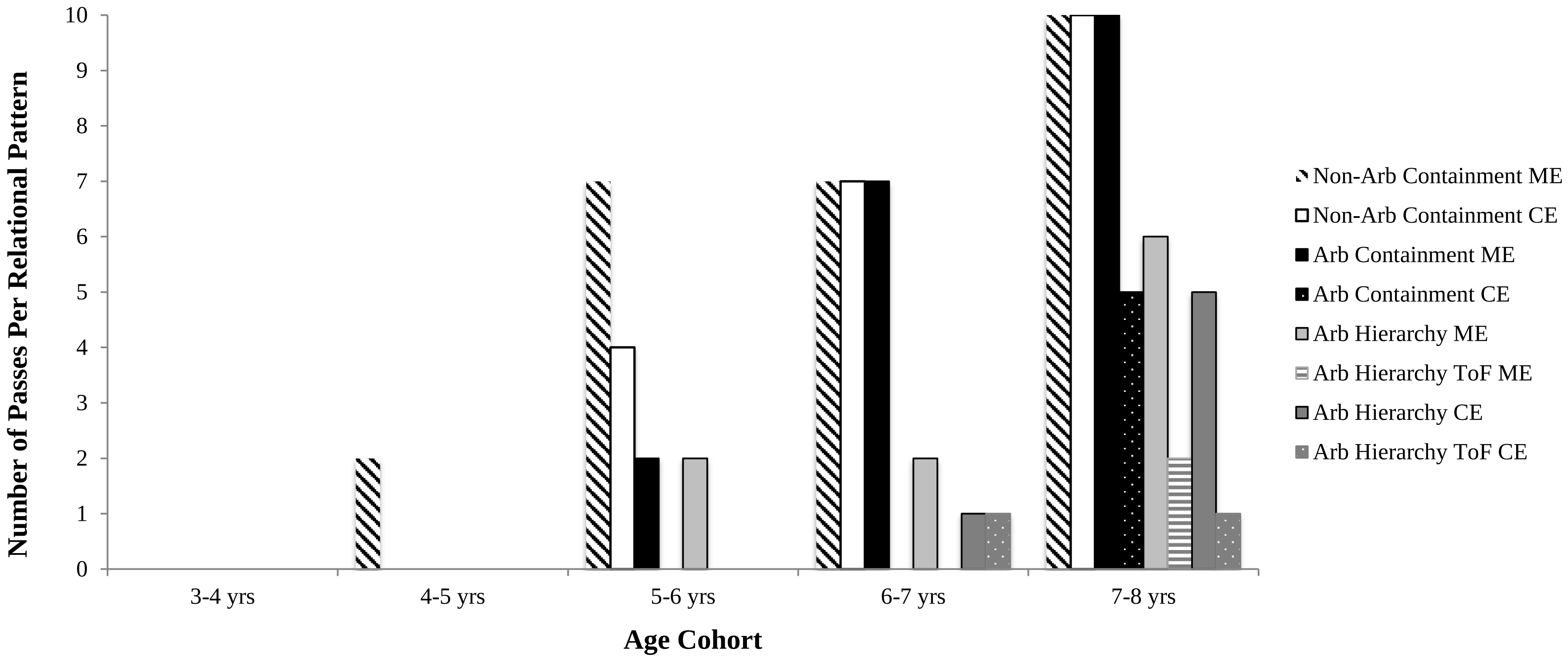


Table 3.7. Test-retest correlations for relational framing test and subtests.

Relational Test	Spearman's Rho correlation
Non-Arbitrary Containment	
Mutual Entailment	.981*
Combinatorial Entailment	.981*
Total Score	.991*
Arbitrary Containment	
Mutual Entailment	.974*
Combinatorial Entailment	.966*
Total Score	.989*
Arbitrary Hierarchy	
Mutual Entailment	.982*
ToF of Mutually Entailed Relations	.803*
Combinatorial Entailment	.946*
ToF of Combinatorially Entailed Relations	.821*
Total Score	.993*
Total Relational Framing	.993*

Note. All scores were raw (unstandardised) scores. * $p < .001$; ToF = Transformation of Function.

Relational Framing Test-Retest Scores

Table 3.7 shows correlations between scores for the initial test and the re-test in the case of all sub-sections of each relation, for each relational total and for the overall relational framing test total. All the tests showed very highly significant test-retest correlations.

Discussion

The current study aimed to investigate relational framing related to hierarchical classification in children. To this end, it involved administering to 50 children ranging from 3 to 8 years of age a custom-made protocol designed to assess relational framing presumed to be relevant to hierarchical classification from relatively simple and concrete up to relatively more complex and abstract – specifically, non-arbitrary containment, arbitrary containment and arbitrary hierarchy, each assessed through testing of ME and CE, as well as, in the case of arbitrary hierarchy, through testing of ToF. This was done with a view to gathering information regarding typical acquisition of relational framing repertoires relevant to classification; furthermore, the protocol was administered more than once to check for reliability. All the children were also tested using a number of other measures including (a) of generalised intellectual (cognitive and linguistic) skill (i.e., the SB5 and the PPVT-4) and (b) of classification skill specifically (i.e., the CCT and the Piagetian Class Inclusion Test). The tests of generalised intellectual performance were administered chiefly to examine the correlation of relational framing of categorisation with intellectual performance more generally, whereas the tests of categorisation were administered to probe the validity of the claim that the relational framing repertoires examined herein are indeed relevant to classification responding.

The data showed first that there was a developmental trend in terms of the acquisition of the overall relational framing repertoire involved, as well as with respect to each of the more specific relational repertoires. This is evident both in the correlational data (Table 3.4) as shown by the high correlation between age and each of the repertoires, and in the descriptive data (Table 3.5) and the data showing the number of passes (Table 3.6) for each cohort for various patterns of relational framing (frames and properties). These patterns suggest that, as might have been predicted, relational framing of categorisation is acquired gradually over the course of several years. The youngest age group (3-4 years) shows little or no capacity in any of the three specific relational repertoires assessed, including even non-arbitrary containment, whereas data for the oldest age group (7-8 years) indicate at least some emergence on each of the three repertoires. All members of this oldest group passed all sections of non-arbitrary containment. However, their performance on the two arbitrary sections and particularly the arbitrary hierarchy section was relatively poor, even though they performed better than the younger groups. These data suggest that children of all age groups might potentially benefit from training in relational frames of categorisation and, given the correlations seen between these patterns of framing and performance on standardised measures of language and cognition, that such training might be intellectually beneficial. The data for number of passes also show that the order of sequencing of relational framing repertoires in the current protocol (i.e., non-arbitrary containment, arbitrary containment and arbitrary hierarchy) was correct and that this is indeed the order of acquisition and would thus also be an appropriate order for assessment and training. Finally, scores on the second administration of the protocol also correlated very highly with scores for the first administration (Table 3.7), thus suggesting the reliability of the protocol.

Regarding the data showing correlations between relational framing and performance on other measures, a number of points in particular are important. First, as previously mentioned, there were high levels of correlation between relational framing and performance on standardised intellectual measures, including both cognitive and linguistic. These correlations were seen both for overall relational framing capacity as well as for specific relational repertoires. These findings further extend a pattern seen in previous RFT studies showing strong correlations between relational framing and intellectual performance (e.g., Moran, Walsh, Stewart, McElwee & Ming, 2015; O'Toole et al., 2009; Ruiz & Luciano, 2011). Second, although there were strong correlations seen between scores for the relational framing repertoires and those for the standardised tests of categorisation specifically, these correlations were not as high as those found for the broad measures of intellectual potential.

Assuming that the relational framing test is tapping into categorisation as a component of intellectual potential more broadly then we might expect that this test would correlate well with both tests of broad intellectual ability as well as with specific tests of categorisation. One circumstance under which the relational framing test might correlate better with a broad test of intellectual ability than with a particular test of categorisation might be if the latter was a particularly specific test of categorisation. The relational framing test might correlate to a reasonable level with such a specific test; however, being quite a broad measure itself, it might correlate still better with a broad test of intelligence or language. This is possibly the pattern that pertains in the case of the class inclusion test, since, while this is a test of categorisation, it assesses a very specific skill. Further support for this explanation is that the class inclusion test does not correlate very highly with the standardised measures of language or intellectual potential but it does still correlate well with the relational framing protocol, and in fact better than with any other measure. In the case of the CCT, however, quite a different pattern is seen. This protocol does correlate relatively well with the relational framing assessment, but, in contrast with the class inclusion test, it also correlates very highly with the standardised measures (in fact, even more so than with the relational framing test) while, in contrast with the relational framing test, showing a particularly low level of correlation with the class inclusion measure. This pattern would suggest that the CCT, despite its name, might in fact not function as much as a test of categorisation in particular as a measure of intellectual functioning more generally. This verdict might seem to be borne out to some extent by consideration of the nature of the testing items, which necessitate responding in accordance with a variety of different non-arbitrary relations between objects at various levels of complexity while not seeming to focus in particular on the assessment of categorisation skill per se. Tests such as this might well tap aspects of categorisation while testing other types of relational repertoire but perhaps little more than other tests of general intellectual ability.

An original version of the current protocol assessed transformation of function not only for arbitrary hierarchy but also for all three framing patterns; indeed, data on transformation of function for all three frames were collected for the current cohort. However, the items assessing transformation of function in the case of the two containment relational patterns were problematic, and thus the data were omitted. Hence, the data reported herein for these two frames focus just on derived mutual and combinatorial entailment alone and do not assess transformation of function. Despite this, the correlations between the overall relational framing protocol and the alternative measures used were strong. This is perhaps not surprising, as (a) transformation of function data *were* collected for arbitrary hierarchy, which might be argued to be particularly important in terms of the repertoire of classification, and (b) RFT protocols that focus on derived relations alone have previously proven to be excellent predictors of intellectual performance (see, e.g., Cassidy, Roche, & Hayes, 2011). Therefore, the data provided for the containment frames would still have been useful in this respect. At the same time, from an RFT point of view, transformation of function is a key property of relational framing, and thus collecting it for all three frames would likely boost predictive power. Engaging in the latter as an extension of the current study would be a useful future research direction.

Another possible critique of the categorisation framing protocol used in the current study might be that it did not provide as thorough or comprehensive a test of arbitrary hierarchical framing as might be desirable. For example, although framing appropriately in response to the cues of *type of* and *contains* is an important part of this pattern, there are other aspects of this pattern that might also be tested. For instance, when two stimuli are framed as being part of the same class, then they should also both be framed as being different from other stimuli framed within a different class and equivalent to each other, in at least some contexts, independent of their physical properties. Such a pattern of responding might be expected from someone with a sufficiently advanced repertoire of hierarchical classification. The current protocol does not involve testing for such relations. Adjusting it so as to do so might be expected to improve its reliability and validity as a test of classification. This would be one useful direction for future work.

Another broader issue for consideration is that when using a relational skills test as a proxy for other skills (i.e., such as categorisation) by way of identifying which skills deficits may need ameliorating in a relational skill training intervention, it is assumed that the relational skills levels are distributed more or less normally within each age cohort. In other words, if one does not know how skills are distributed across larger samples, one cannot know the significance of a relational test score that is deviating from the mean. Analysis of this kind would not have been possible with the current data set, as it is too small, but this may be an important consideration for researchers interested in using relational tests as either proxies for other intellectual skills or as screening tests to assess deficits prior to intervention.

Another relevant direction for future studies might be the exploration of the capacity for establishing and/or strengthening repertoires of classification framing. If the current protocol (or a refined version of it) is seen as providing a reasonable measure of relational framing relevant to classification, as seems possible based on these data, then in future studies, it might be used to identify deficits in this repertoire as a precursor to the training of the relevant relational skills. An intervention for one or more aspects of classification framing might be expected to boost not only this repertoire itself but also—analogous to the outcomes seen in the case of other relational frame training interventions—intellectual performance more generally. Indeed, given the centrality of classification to the average person’s intellectual repertoire, perhaps such training might have a more substantial effect in this regard than training involving other frames.

One further possible direction for future research might be the examination of the relationship and/or interaction between classification framing and other framing repertoires. Such work might investigate various phenomena, including correlation between classification and various other patterns of framing, as well as the effect of training particular frames on performance in accordance with classification framing. For example, as suggested previously, sameness and difference frames may be particularly important precursors and/or accompaniments of hierarchical framing, suggesting that training fluency and/or flexibility on these frames might be particularly beneficial with respect to the acquisition and/or strengthening of hierarchical framing and categorisation framing more broadly.

In summary, this is a promising initial result for the assessment of classification framing. The pattern of data acquired from the protocol used to measure this repertoire in combination with that from a suite of other measures of intellectual functioning, including ones focused on classification per se, might suggest that the former is indeed tapping into important aspects of categorisation. This extends research on relational framing in general and provides an impetus for further work into classification per se from an RFT point of view. Further research might usefully extend the current study by incorporating additional testing of features of categorisation as a pattern of responding as well as by facilitating the training of this repertoire in typically developing children or children with developmental delays who show relevant deficits.

Chapter 4

Studies 2 and 3

Training Arbitrary Containment Repertoires in Young Children*

* Portions of this chapter have been submitted for publication:

Mulhern, T., Stewart, I., & McElwee, J. (under submission). Facilitating relational framing of classification in young children.

As described in Chapter 2, derived relational responding or relational framing repertoires have been positively correlated with a broad spectrum of cognitive abilities, including language, reasoning and problem solving (Dymond & Roche, 2013). For example, Galizio and colleagues (2001) argued that derived equivalence is similar to the process of category clustering. Galizio and colleagues further argued that a behaviour analytic conceptualisation of categorisation may significantly contribute to our understanding of this complex phenomenon. The previously outlined research indicates that there are two fundamental (empirically supported) arguments made by RFT in regards to the acquisition of relational framing repertoires and educational repertoires in general. The first is that verbal relational skills offer the foundation for a broad spectrum of cognitive abilities that are related to educational achievement (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000). The second is that multiple exemplar training (MET) is a necessary component of procedures designed to establish flexible cognitive skills (Barnes-Holmes, Barnes-Holmes, & McHugh, 2004).

Empirical evidence previously cited suggests that RFT may provide an effective behavioural and functional approach to the comprehension, assessment, and training of a broad spectrum of verbal and cognitive events. The current study aims to extend the research in the domains of RFT and behaviour analysis more generally by assessing and increasing arbitrary containment relational repertoires in typically developing young children.

As suggested in Chapters 2 and 3, arbitrary containment repertoires are theorised to form a cornerstone for categorisation and hierarchical classification repertoires (Slattery & Stewart, 2014). In the current chapter, two studies (2 and 3) are presented which focus on the training of these relations. Study 2 is a pilot study, which aimed to augment arbitrary containment relational repertoires in a 5-year-old child in the context of a multiple baseline across responses design. Study 3 extended the work of Study 2 by employing the same training procedure within the context of a combined multiple baseline design across responses and across participants (similar to that used in Berens and Hayes, 2007, for example) with three 5-year-old children. Study 3 also included the use of a comparison group that did not receive arbitrary containment training in order to help assess the impact of arbitrary containment training on untrained variables, including class inclusion responding and standardised measures of language and categorisation.

Study 2 Method

Participant & Setting

Prior ethical approval for the recruitment of participants for this study was obtained from the research ethics committee of the National University of Ireland, Galway. A typically developing 5-year 4-month-old Irish female participated in the pilot study (Pt. 2.1). Consent for conducting the study was obtained from the child's parents and verbal consent was also obtained from the participant.

The participant was selected based on her performance on an assessment of non-arbitrary and arbitrary containment relational responding (see Materials) in which she passed the former (by achieving a score of 80% or more) and failed the latter (by achieving a score of 60%) or less. The score of 80% or more was selected based upon the results of Study 1, which indicated that a strong repertoire of non-arbitrary containment was present if a participant emitted correct responses for 80% or more of the trials. The score of 60% or less of correct arbitrary containment responses was selected as the results of Study 1 indicated that this rate of responding (or lower) was indicative of a weak and unestablished repertoire that was no greater than chance responding.

Baseline, training, generalisation and maintenance sessions were conducted within the participants' own home within a quiet room. Sessions occurred between three and five times per week for 40-minute sessions.

Materials

The participant was exposed to two different tests of relational responding, namely, non-arbitrary containment and arbitrary containment.

Non-Arbitrary Containment Assessment: The materials included 3-D square boxes in three different sizes (large, medium, or small) and a range of colours (red, yellow, green, blue, purple, orange, black, white, and pink).

Arbitrary Containment Assessment: The materials included four stimulus sets of nonsense syllables (see Table 4.1), which were presented as textual stimuli on a laptop screen using PowerPoint. The Participant was assessed both before and after training (both immediately as well as in maintenance testing) for all four stimulus sets (i.e., Sets 1 – 4). The participant was trained using Stimulus Set 1 alone.

Table 4.1. Arbitrary Containment Stimulus Sets.

	Stimulus Set 1	**Stimulus Set 2**	**Stimulus Set 3**	**Stimulus Set 4**
A	Blorg	Zimp	Quig	Stak
B	Grap	Crint	Donk	Timp
C	Plak	Wark	Frik	Vink

Experimental Design

The current study employed a multiple baseline design across responses. Multiple baseline design across responses (also known as multiple baseline design across behaviours) offer a number of advantages. Firstly, this design can increase the internal validity of the findings generated by the research by providing the experimenter with a means to assess and demonstrate intra-subject direct replication. Secondly, this design does not require a return to baseline conditions in order to assess experimental control, thereby eliminating the practical and ethical issues associated with withdrawal designs. Thirdly, such a design provides a practitioner with a practical method of assessing non-reversible functional skills (e.g., greeting responses, spelling, walking appropriately, etc.) that are difficult to establish and are also inappropriate or unethical to reverse. Finally, as behaviour maintenance is of interest to both behaviour analysts and educators, this design provides a suitable paradigm for repeatedly examining progress over time (Gast, Lloyd & Ledford, 2014).

Within the context of the multiple baseline design across responses, the first of these responses, mutual entailment of arbitrary containment relational responding, was assessed until a stable level of responding was observed (meaning that scores for the last three probes were equal or that the majority of scores among the last five probes were equal within baseline sessions. Following this, the participant was entered into Phase 1 of training, while the remaining responses (i.e., transformation of stimulus function across two stimuli, combinatorial entailment, and transformation of stimulus function across three stimuli) continued to be assessed within baseline conditions. During each phase of training, the remaining untrained relational components were probed during experimental sessions. This sequence of testing was repeated for (a) transformation of stimulus function of mutually entailed relations, (b) combinatorial entailment, and (c) transformation of stimulus function of combinatorially entailed relations in which the participant was exposed to seven, ten, and twelve baseline sessions respectively.

Procedure

The participant was assessed for non-arbitrary and arbitrary containment relations over two sessions, each of which lasted roughly 40 minutes.

Assessing Non-Arbitrary Containment. The participant was assessed for non-arbitrary containment relations (see Table 4.2 for non-arbitrary containment trial types) using differently sized and coloured boxes. She was assessed for (a) mutual entailment, (b) transformation of function via mutual entailment, (c) combinatorial entailment, and (d) transformation of stimulus functions via combinatorial entailment.

On each trial, the participant was first shown the relevant stimuli, the experimenter then described the relationship between the stimuli and further demonstrated this relationship by manipulating the stimuli (e.g., “A red box is inside the blue box” – the experimenter would then place the red box inside the blue box). The participant was then asked a question concerning the relationship between the stimuli, which required her to answer either “yes” or “no” (e.g., “Does the blue box contain a red box?”). No feedback was provided regarding performance and no reinforcement was provided for correct or incorrect responding.

Table 4.2. Trial types used to assess non-arbitrary containment repertoires.

Property	Statement	Question [Answer]
Mutual Entailment (2 stimuli)	The [A] box is inside the [B] box	Is the [A] box inside the [B] box? [Yes]
	The [A] box is inside the [B] box	Is the [B] box inside the [A] box? [No]
	The [A] box is inside the [B] box	Does the [A] box contain the [B] box? [No]
	The [A] box is inside the [B] box	Does the [B] box contain the [A] box? [Yes]
	The [B] box is inside the [A] box	Is the [A] box inside the [B] box? [No]
	The [B] box is inside the [A] box	Is the [B] box inside the [A] box? [Yes]
	The [B] box is inside the [A] box	Does the [A] box contain the [B] box? [Yes]
	The [B] box is inside the [A] box	Does the [B] box contain the [A] box? [No]
	The [A] box contains the [B] box	Is the [A] box inside the [B] box? [No]
	The [A] box contains the [B] box	Is the [B] box inside the [A] box? [Yes]
	The [A] box contains the [B] box	Does the [A] box contain the [B] box? [Yes]
	The [A] box contains the [B] box	Does the [B] box contain the [A] box? [No]
	The [B] box contains the [A] box	Is the [A] box inside the [B] box? [Yes]
	The [B] box contains the [A] box	Is the [B] box inside the [A] box? [No]
	The [B] box contains the [A] box.	Does the [A] box contain the [B] box? [No]
	The [B] box contains the [A] box	Does the [B] box contain the [A] box? [Yes]
ToF via Mutual Entailment (2 stimuli)	The [A] box is inside the [B] box. Brian likes the [A] box, and Sarah likes the [B] box.	Is there a box that Brian likes inside box [B]? [Yes]
	The [A] box is inside the [B] box. Brian likes the [A] box, and Sarah likes the [B] box.	Is there a box that Sarah likes inside box [A]? [No]
	The [A] box is inside the [B] box. Brian likes the [A] box, and Sarah likes the [B] box.	Does box [A] contain a box that Sarah likes? [No]
	The [A] box is inside the [B] box. Brian likes the [A] box, and Sarah likes the [B] box.	Does box [B] contain a box that Brian likes? [Yes]
	The [B] box is inside the [A] box. Brian likes the [A] box, and Sarah likes the [B] box.	Is there a box that Brian likes inside box [B]? [No]
	The [B] box is inside the [A] box. Brian likes the [A] box, and Sarah likes the [B] box.	Is there a box that Sarah likes inside box [A]? [Yes]
	The [B] box is inside the [A] box. Brian likes the [A] box, and Sarah likes the [B] box.	Does box [A] contain a box that Sarah likes? [Yes]
	The [B] box is inside the [A] box. Brian likes the [A] box, and Sarah likes the [B] box.	Does box [B] contain a box that Brian likes? [No]
	Box [A] contains box [B]. Brian likes box [A] and Sarah likes box [B].	Is there a box that Brian likes inside box [B]? [No]
	Box [A] contains box [B]. Brian likes box [A] and Sarah likes box [B].	Is there a box that Sarah likes inside box [A]? [Yes]
	Box [A] contains box [B]. Brian likes box [A] and Sarah likes box [B].	Does box [A] contain a box that Sarah likes? [Yes]
	Box [A] contains box [B]. Brian likes box [A] and Sarah likes box [B].	Does box [B] contain a box that Brian likes? [No]
	Box [B] contains box [A]. Brian likes box [A] and Sarah likes box [B].	Is there a box that Brian likes inside box [B]? [Yes]
	Box [B] contains box [A]. Brian likes box [A] and Sarah likes box [B].	Is there a box that Sarah likes inside box [A]? [No]
	Box [B] contains box [A]. Brian likes box [A] and Sarah likes box [B].	Does box [A] contain a box that Sarah likes? [No]
	Box [B] contains box [A]. Brian likes box [A] and Sarah likes box [B].	Does box [B] contain a box that Brian likes? [Yes]
Comb. Entailment (3 stimuli)	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Is the [A] box inside the [C] box? [Yes]
	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Is the [C] box inside the [A] box? [No]
	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Does the [A] box contain the [C] box? [No]
	The [A] box is inside the [B] box; The [B] box is inside the [C] box	Does the [C] box contain the [A] box? [Yes]
	The [C] box is inside the [B] box; The [B] box is inside the [A] box	Is the [A] box inside the [C] box? [No]
	The [C] box is inside the [B] box The [B] box is inside the [A] box	Is the [C] box inside the [A] box? [Yes]

	The [C] box is inside the [B] box; The [B] box is inside the [A] box	Does the [A] box contain the [C] box? [Yes]
	The [C] box is inside the [B] box; The [B] box is inside the [A] box	Does the [C] box contain the [A] box? [No]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Is the [A] box inside the [C] box? [No]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Is the [C] box inside the [A] box? [Yes]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Does the [A] box contain the [C] box? [Yes]
	The [A] box contains the [B] box; The [B] box contains the [C] box	Does the [C] box contain the [A] box? [No]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Is the [A] box inside the [C] box? [Yes]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Is the [C] box inside the [A] box? [No]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Does the [A] box contain the [C] box? [No]
	The [C] box contains the [B] box; The [B] box contains the [A] box	Does the [C] box contain the [A] box? [Yes]
ToF via Comb. Entailment (3 stimuli)	Box [A] is inside box [B], box [B] is inside box [C]. George likes box [A], and Rebecca likes box [C].	Is there a box that George likes inside box [C]? [Yes]
	Box [A] is inside box [B], box [B] is inside box [C]. George likes box [A], and Rebecca likes box [C].	Is there a box that Rebecca likes inside box [A]? [No]
	Box [A] is inside box [B], box [B] is inside box [C]. George likes box [A], and Rebecca likes box [C].	Does box [A] contain a box that Rebecca likes? [No]
	Box [A] is inside box [B], box [B] is inside box [C]. George likes box [A], and Rebecca likes box [C].	Does box [C] contain a box that George likes? [Yes]
	Box [C] is inside box [B], box [B] is inside box [A]. George likes box [A] and Rebecca likes box [C].	Is there a box that George likes inside box [C]? [No]
	Box [C] is inside box [B], box [B] is inside box [A]. George likes box [A] and Rebecca likes box [C].	Is there a box that Rebecca likes inside box [A]? [Yes]
	Box [C] is inside box [B], box [B] is inside box [A]. George likes box [A] and Rebecca likes box [C].	Does box [A] contain a box that Rebecca likes? [Yes]
	Box [C] is inside box [B], box [B] is inside box [A]. George likes box [A] and Rebecca likes box [C].	Does box [C] contain a box that George likes? [No]
	Box [A] contains box [B], box [B] contains box [C]. George likes box [A], and Rebecca likes box [C].	Is there a box that George likes inside box [C]? [No]
	Box [A] contains box [B], box [B] contains box [C]. George likes box [A], and Rebecca likes box [C].	Is there a box that Rebecca likes inside box [A]? [Yes]
	Box [A] contains box [B], box [B] contains box [C]. George likes box [A], and Rebecca likes box [C].	Does box [A] contain a box that Rebecca likes? [Yes]
	Box [A] contains box [B], box [B] contains box [C]. George likes box [A], and Rebecca likes box [C].	Does box [C] contain a box that George likes? [No]
	Box [C] contains box [B], box [B] contains box [A]. George likes box [A], and Rebecca likes box [C].	Is there a box that George likes inside box [C]? [Yes]
	Box [C] contains box [B], box [B] contains box [A]. George likes box [A], and Rebecca likes box [C].	Is there a box that Rebecca likes inside box [A]? [No]
	Box [C] contains box [B], box [B] contains box [A]. George likes box [A], and Rebecca likes box [C].	Does box [A] contain a box that Rebecca likes? [No]
	Box [C] contains box [B], box [B] contains box [A]. George likes box [A], and Rebecca likes box [C].	Does box [C] contain a box that George likes? [Yes]

Note. ToF = Transformation of Function; The codes [A], [B] and [C] signify different colours within each trial. The actual colours of the box stimuli (e.g., red, green, blue, yellow, pink, purple etc.) varied quasi-randomly across trials.

In trials for mutual entailment, the participant was shown two boxes (one inside another) and all questions within this phase of assessment focused on the relationship between these two stimuli. A total of 16 such trials were presented in random order. Trials for the transformation of stimulus function via mutual entailment were similar except for the addition of information regarding stimulus function (e.g., “Brian likes the colour blue”). These questions differed from the questions for mutual entailment itself in that they focused on the potential change in stimulus functions of one of the two stimuli in the mutually entailed relation by virtue of being in that relationship (e.g., “Is the red box inside a box that Brian likes?”). This section too was assessed across 16 questions presented in random order.

In trials for combinatorial entailment, the participants were presented with three boxes (the first inside the second and the second inside the third) and all 16 questions within this phase, which were presented in random order, focused on the derivation of relations between stimuli A and C. Trials for the transformation of stimulus function via combinatorial entailment were similar except for the addition of information regarding stimulus function (e.g., “Brian likes the colour blue), analogous to the case for mutual entailment. This section too was assessed across 16 questions presented in random order. The criterion for passing this assessment was showing at least 80% (i.e., approximately 52 out of 64) correct.

Assessing Arbitrary Containment. The participant was assessed for arbitrary containment relations using a laptop-based PowerPoint presentation involving nonsense syllables (see Table 4.3 for arbitrary containment trial types). The participant was assessed for (a) mutual entailment, (b) transformation of function via mutual entailment, (c) combinatorial entailment, and (d) transformation of stimulus functions via combinatorial entailment.

On each trial, the participant was first presented with text on a laptop screen that described the arbitrary containment relation between two nonsense syllables; for example "The Blorg is inside the Grap" (see Table 4.1 for the sets of stimuli used in the relations in this phase). The experimenter read the text aloud to the participant and then asked them a question based on the stimuli involved which could be answered in the form of a "yes" or "no" response (e.g., "Does the Grap contain the Blorg?"). No feedback was provided regarding performance and no reinforcement was provided for correct or incorrect responding.

In trials for mutual entailment, the relationship between two stimuli was outlined to the participant and she was required to answer a subsequent question with either a "yes" or "no" based on the presented relation. A total of sixteen trials (see Table 3.3) were presented in random order. Trials for the transformation of stimulus function via mutual entailment were similar except for the addition of information regarding stimulus functions (See Table 3.3). These questions focused on the potential change in stimulus functions of one of the two stimuli in the mutually entailed relation by virtue of being in that relationship. This section too was assessed across 16 questions presented in random order.

In trials for combinatorial entailment, the relationship between three stimuli was outlined and the participant was required to answer a subsequent question with either "yes" or "no" based on the presented relation. A total of sixteen trials (see Table 3.3) were presented in random order.

Table 4.3. Trial types used to assess and train arbitrary containment.

Property	Statement	Question [Answer]
Mutual Entailment (2 stimuli)	[A] is inside [B]	Is [A] inside [B]? [Yes]
	[A] is inside [B]	Is [B] inside [A]? [No]
	[A] is inside [B]	Does [A] contain [B]? [No]
	[A] is inside [B]	Does [B] contain [A]? [Yes]
	[A] contains [B]	Is [A] inside [B]? [No]
	[A] contains [B]	Is [B] inside [A]? [Yes]
	[A] contains [B]	Does [A] contain [B]? [Yes]
	[A] contains [B]	Does [B] contain [A]? [No]
	[B] is inside [A]	Is [A] inside [B]? [No]
	[B] is inside [A]	Is [B] inside [A]? [Yes]
	[B] is inside [A]	Does [A] contain [B]? [Yes]
	[B] is inside [A]	Does [B] contain [A]? [No]
	[B] contains [A]	Is [A] inside [B]? [Yes]
	[B] contains [A]	Is [B] inside [A]? [No]
	[B] contains [A]	Does [A] contain [B]? [No]
	[B] contains [A]	Does [B] contain [A]? [Yes]
Transformation of Function (2 stimuli)	[A] is inside [B]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property A] inside [B]? [Yes]
	[A] is inside [B]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property B] inside [A]? [No]
	[A] is inside [B]. [A] has [Property A], and [B] has [Property B]	Does [A] contain something with [Property B]? [No]
	[A] is inside [B]. [A] has [Property A], and [B] has [Property B]	Does [B] contain something with [Property A]? [Yes]
	[A] contains [B]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property A] inside [B]? [No]
	[A] contains [B]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property B] inside [A]? [Yes]
	[A] contains [B]. [A] has [Property A], and [B] has [Property B]	Does [A] contain something with [Property B]? [Yes]
	[A] contains [B]. [A] has [Property A], and [B] has [Property B]	Does [B] contain something with [Property A]? [No]
	[B] is inside [A]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property A] inside [B]? [No]
	[B] is inside [A]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property B] inside [A]? [Yes]
	[B] is inside [A]. [A] has [Property A], and [B] has [Property B]	Does [A] contain something with [Property B]? [Yes]
	[B] is inside [A]. [A] has [Property A], and [B] has [Property B]	Does [B] contain something with [Property A]? [No]
	[B] contains [A]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property A] inside [B]? [Yes]
	[B] contains [A]. [A] has [Property A], and [B] has [Property B]	Is there something with [Property B] inside [A]? [No]
	[B] contains [A]. [A] has [Property A], and [B] has [Property B]	Does [A] contain something with [Property B]? [No]
	[B] contains [A]. [A] has [Property A], and [B] has [Property B]	Does [B] contain something with [Property A]? [Yes]
Combinatorial Entailment (3 stimuli)	[A] is inside [B], [B] is inside [C]	Is [A] inside [C]? [Yes]
	[A] is inside [B], [B] is inside [C]	Is [C] inside [A]? [No]
	[A] is inside [B], [B] is inside [C]	Does [A] contain [C]? [No]

	[A] is inside [B], [B] is inside [C]	Does [C] contain [A]? [Yes]
	[A] contains [B], [B] contains [C]	Is [A] inside [C]? [No]
	[A] contains [B], [B] contains [C]	Is [C] inside [A]? [Yes]
	[A] contains [B], [B] contains [C]	Does [A] contain [C]? [Yes]
	[A] contains [B], [B] contains [C]	Does [C] contain [A]? [No]
	[C] is inside [B], [B] is inside [A]	Is [A] inside [C]? [No]
	[C] is inside [B], [B] is inside [A]	Is [C] inside [A]? [Yes]
	[C] is inside [B], [B] is inside [A]	Does [A] contain [C]? [Yes]
	[C] is inside [B], [B] is inside [A]	Does [C] contain [A]? [No]
	[C] contains [B], [B] contains [A]	Is [A] inside [C]? [Yes]
	[C] contains [B], [B] contains [A]	Is [C] inside [A]? [No]
	[C] contains [B], [B] contains [A]	Does [A] contain [C]? [No]
	[C] contains [B], [B] contains [A]	Does [C] contain [A]? [Yes]
Transformation of Function (3 Stimuli)	[A] is inside [B], [B] is inside [C]. [A] has [Property A], and [C] has [Property C]	Is there something with [Property A] inside [C]? [Yes]
	[A] is inside [B], [B] is inside [C]. [A] has [Property A], and [C] has [Property C]	Is there something with [Property C] inside [A]? [No]
	[A] is inside [B], [B] is inside [C]. [A] has [Property A], and [C] has [Property C]	Does [A] contain something with [Property C]? [No]
	[A] is inside [B], [B] is inside [C]. [A] has [Property A], and [C] has [Property C]	Does [C] contain something with [Property A]? [Yes]
	[A] contains [B], [B] contains [C]. [A] has [Property A], and [C] has [Property C].	Is there something with [Property A] inside [C]? [No]
	[A] contains [B], [B] contains [C]. [A] has [Property A], and [C] has [Property C].	Is there something with [Property C] inside [A]? [Yes]
	[A] contains [B], [B] contains [C]. [A] has [Property A], and [C] has [Property C].	Does [A] contain something with [Property C]? [Yes]
	[A] contains [B], [B] contains [C]. [A] has [Property A], and [C] has [Property C].	Does [C] contain something with [Property A]? [No]
	[C] is inside [B], [B] is inside [A]. [A] has [Property A], and [C] has [Property C]	Is there something with [Property A] inside [C]? [No]
	[C] is inside [B], [B] is inside [A]. [A] has [Property A], and [C] has [Property C]	Is there something with [Property C] inside [A]? [Yes]
	[C] is inside [B], [B] is inside [A]. [A] has [Property A], and [C] has [Property C]	Does [A] contain something with [Property C]? [Yes]
	[C] is inside [B], [B] is inside [A]. [A] has [Property A], and [C] has [Property C]	Does [C] contain something with [Property A]? [No]
	[C] contains [B], [B] contains [A]. [A] has [Property A], and [C] has [Property C]	Is there something with [Property A] inside [C]? [Yes]
	[C] contains [B], [B] contains [A]. [A] has [Property A], and [C] has [Property C]	Is there something with [Property C] inside [A]? [No]
	[C] contains [B], [B] contains [A]. [A] has [Property A], and [C] has [Property C]	Does [A] contain something with [Property C]? [No]
	[C] contains [B], [B] contains [A]. [A] has [Property A], and [C] has [Property C]	Does [C] contain something with [Property A]? [Yes]

Note. The codes [A], [B] and [C] signify different nonsense syllables within each trial. The actual nonsense stimuli (e.g., blorg, grap, plak, zimp, crint, wark etc.) varied quasi-randomly across trials.

Trials for the transformation of stimulus function via combinatorial entailment were similar except for the addition of information regarding stimulus function (See Table 3.3). These questions focused on the potential change in stimulus functions of one of the two stimuli in the combinatorially entailed relation by virtue of being in that relationship. This section too was assessed across 16 questions presented in random order. The criterion for passing this assessment was showing at least 80% (i.e., approximately 52 out of 64) correct.

Training Arbitrary Containment. Pt. 2.1 was exposed to arbitrary containment relational training over the course of two to three weeks. Each training session lasted approximately 40 minutes and sessions were held five times per week.

As explained in the experimental design section, the participant was exposed to baseline session before receiving training in accordance with the sequential logic of the multiple baseline design. The baseline assessment was identical to the assessment of arbitrary containment relations already described.

Training was presented in four phases corresponding to each of the four components of the relational operant outlined previously in the assessment phase (i.e., mutual entailment, transformation of function via mutual entailment, combinatorial entailment, and transformation of function via combinatorial entailment). Within each phase the remaining untrained relational components were probed (using the same fixed number of trials without feedback or reinforcement as had been used in the assessment phase) during experimental sessions. Training in each phase was the same as assessment of that phase except for the provision of reinforcement and feedback as appropriate and except that trials continued to be presented within a phase until pass criteria for that phase were met (see Appendix B for training protocols and assessments).

During training sessions, correct responses were reinforced on an FR1 schedule of specific praise relating to the task (e.g., “That’s right! A is inside B”), in addition to an FR1 schedule of tokens (laminated and velcroed pictures which could then be affixed onto a token board), which were then exchanged on an FR4 schedule for a self-chosen sticker. Following incorrect responses the experimenter would say “No, that’s not it”, and provide contingent feedback (e.g., “B *isn’t* inside A, because A is inside B!”) and the trial would then be represented to the participant until they emitted a correct response. Apart from this, the participant was also told that meeting the goals of each session would result in an additional prize (e.g., rubbers, pencils, and markers which she could for herself from a box). The goal for the initial training session on all training phases was 50% or more correct. The subsequent goal for each training session was for the participant to beat her previous score by at least one more correct response than the previous training session. The participant was told before each training session how many correct responses she need to “beat” her previous score and receive the additional prize at the end of the session. She was also informed that if she did not meet this goal, she would not earn a prize for that session. This contingency remained in place until the participant met criterion (90% across two successive sessions or 100% correct within one session), which signalled the start of the next phase of training.

Generalisation Assessment. Following the completion of the each phase of experimental training, the participant was exposed to generalisation probes using Stimulus Sets 2. 3 and 4. For instance, upon the completion of Phase 1 of training, the participant was then assessed for generalisation of mutually entailed relations using Stimulus Sets 2, 3 and 4. Generalisation probes were identical to that of baseline assessments and were also comprised of 64 trials. No feedback was provided in these trials. Criterion for demonstration of generalisation was 90% correct. The rate of 90% was selected as the research generated from Study 1 indicated that a strong arbitrary containment repertoire was present if an individual exhibited 80% or more correct arbitrary containment responses. As such, the rate of 90% was selected in order to ensure that the participant had acquired a strong generalized arbitrary containment repertoire.

Maintenance Assessment. The participant was assessed for maintenance of responding one week, two weeks, three weeks and four weeks after the completion of each phase of training using Stimulus Set 1. For example, after the participant met mastery criterion for Phase 1 of training, she was then assessed for maintenance of responding one week later. Maintenance probes were identical to those of baseline assessments and were comprised of 64 trials. No feedback was provided in these trials. Criterion for demonstration of maintenance was 90% correct. As outlined previously, the rate of 90% was selected based upon the results generated in Study 1 (see above for rationale).

Inter-Observer Agreement. For 59.4% of all baseline trials, 63.6% of all training trials, 100% of all generalisation trials, and 50% of all maintenance trials, a secondary data collector independently scored each trial as correct or incorrect. The secondary data collector was required to reach 100% accuracy on three consecutive mock baseline sessions prior to scoring an experimental session. Agreement data was then collected across all phase types. An agreement was scored when both observers scored a trial as being either correct or incorrect. A disagreement was scored if the observers recorded the trial differently. A percentage agreement score was calculated by dividing the total agreements by the total agreements plus total disagreements and multiplying by 100. This resulted in a total agreement score of 99.74% (98.44% - 100%).

Procedural Integrity. For 59.4% of all baseline trials, 63.6% of all training trials, 100% of all generalisation trials, and 50% of all maintenance trials, a secondary data collector collected procedural integrity data. Two measures of integrity were scored for every trial: trial presentation and correct consequence provided. For each of these categories, either a yes or no was scored. If any item was scored as no, the entire trial was scored as incorrect. The total number of trials scored as correct were divided by the total number of trials scored. This resulted in an integrity score of 99.87% (98.44% - 100%).

Results and Discussion

The initial relational assessment showed that the participant met criterion (i.e., achieved over 80% correct) for the non-arbitrary containment assessment overall (see Table 4.4) while also failing the arbitrary containment assessment (see Table 4.5).

Table 4.4. Percentage of correct non-arbitrary containment relational responses for Pt. 2.1.

Mutual Entailment	ToF Mutual Entailment	Combinatorial Entailment	ToF Combinatorial Entailment	Total Score
100%	100%	100%	100%	100%

Note. ToF = Transformation of stimulus function.

Table 4.5. Percentage of correct arbitrary containment relational responses for Pt. 2.1.

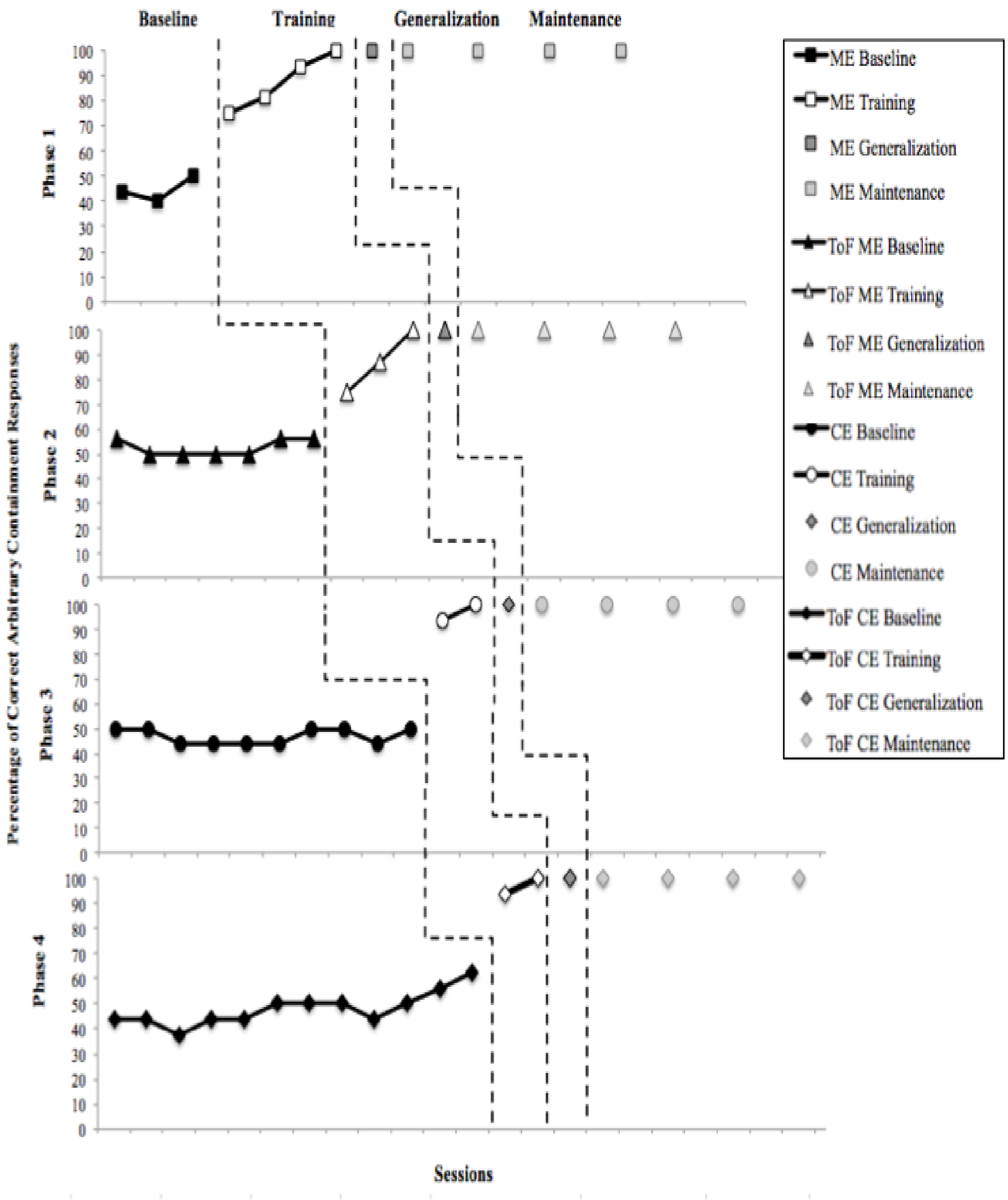
Mutual Entailment	**ToF Mutual Entailment**	**Combinatorial Entailment**	**ToF Combinatorial Entailment**	**Total Score**
43.75%	56.25%	50%	43.75%	48.44%

Note. ToF = Transformation of stimulus function.

The results of the multiple-baseline analysis are displayed in Figure 4.1. Baseline scores for relational components, with the exception of transformation of function of combinatorial entailment, remained stable prior to the introduction of arbitrary containment relational response training. The sequential introduction of arbitrary containment training for each relational component resulted in considerable increases over baseline levels. The majority of these changes were obtained without corresponding changes in the baselines of the remaining untrained relational components, with the exception of transformation of function of combinatorial entailment, which increased minimally following the introduction of combinatorial entailment arbitrary containment training.

The results of the multiple-baseline analysis for generalisation across untrained stimuli and maintenance of training effects are also presented in Figure 4.1. These data demonstrate that she displayed generalisation and maintenance of arbitrary containment relational repertoires.

Figure 4.1. Multiple baseline across components design displaying correct arbitrary containment responses for Pt. 2.1.



Note. ME = Mutual entailment; ToF ME = Transformation of function of mutually entailed relations; CE = Combinatorial entailment; ToF CE = Transformation of function of combinatorially entailed relations.

Pt. 2.1's baseline arbitrary containment mutual entailment scores remained between 43.75% and 50% throughout baseline testing. Following the introduction of arbitrary containment mutual entailment training, her mutual entailment performance increased significantly from baseline levels to 100% over four sessions. The participant was then exposed to generalisation probes across untrained stimuli after training was terminated. The results of the generalisation probe indicated that she had successfully generalised arbitrary containment mutual entailment responses across untrained stimuli. Maintenance probes were then conducted one week, two weeks, three weeks, and four weeks following the termination of training and also indicated that the participants training was successfully maintained for up to a month after training was terminated.

Baseline testing indicated that Pt. 2.1's transformation of function of mutual entailment remained between 50% and 56.25% throughout baseline testing. The participants' performance across this relational component increased significantly from baseline levels to 100% across three sessions when introduced to arbitrary containment transformation of function of mutual entailment training. The results of generalisation probes also indicated that the participant had successfully generalised responding across novel stimuli. The results of maintenance probes also indicated that the participant had successfully maintained training effects for up to four weeks post-training.

Pt. 2.1's baseline arbitrary containment combinatorial entailment scores remained between 43.75% and 50% throughout baseline testing. Her arbitrary containment combinatorial entailment significantly increased from baseline levels to 100% across two sessions once exposed to arbitrary containment combinatorial entailment training. She also demonstrated successful generalisation across untrained stimuli, in addition to maintenance of training effects for up to four weeks following the cessation of training.

Baseline testing indicated Pt. 2.1's transformation of function of combinatorial entailment remained between 37.5% and 50% for the majority of baseline testing. However, her responding began to increase within the final three baseline sessions to 62.5%. Following the introduction of arbitrary containment transformation of function of combinatorial entailment training, her responding increased to 100% across two training sessions. The results of generalisation probes also indicated that she had successfully generalised responding across novel stimuli. The results of maintenance probes also indicated that she had successfully maintained training effects for up to four weeks post-training.

The present study indicates that it is possible to facilitate arbitrary containment relational repertoires with a typically developing child. Given the lack of research within this area, the current study offers further insight into the implementation of successful training procedures to facilitate arbitrary containment repertoires and relational responding as a whole, and also adds to the current literature base within this area. The findings of the current study indicate that the combination of arbitrary containment training, contingent feedback and positive reinforcement procedures comprise an effective training method for successfully establishing arbitrary containment repertoires.

Study 3 Introduction

Study 2 demonstrated successful training and testing of arbitrary containment relational responding in a 5-year-old child. The current study (i.e., Study 3) extended the previous research in a number of ways. Firstly, Study 3 employed an experimental design with greater experimental control (i.e., combined multiple baseline across people and across relational components) than that of Study 2. Secondly, Study 3 provided a more comprehensive maintenance assessment than that of Study 2 by assessing arbitrary containment relational responding with the trained stimulus set (i.e., Stimulus Set 1) and with untrained stimulus sets (i.e., Stimulus Sets 2 – 4), thereby determining whether arbitrary containment training resulted in the maintenance of both trained and generalised responding up to one month following the cessation of training. Finally, Study 3 also assessed participants both pre- and post-training for their performance on measures of language and categorisation and compared the training participants' performance on these measures to participants who did not receive training.

Study 3 Method

Participants & Setting

Prior ethical approval for recruitment of participants for this study was obtained from the research ethics committee of the National University of Ireland, Galway. Six typically developing Irish children in the 5- to 6-year-old range all enrolled in a rural Irish primary school served as participants. Consent for conducting the study was obtained from the principal in the school involved. Parental consent was obtained for each child was participated and verbal consent was also obtained from each of the participants.

As with Study 2, all six participants were selected based on their performance on an assessment of non-arbitrary and arbitrary containment relational responding. Three children (mean age = 5 years 3 months; SD = 1.5 months) were randomly assigned to the training group (i.e., the group that would receive arbitrary containment training). Training participants included Pt. 3.1, Pt. 3.2 and Pt. 3.3. Pt. 3.1 was a 5-year 5-month-old female, Pt. 3.2 was a 5-year 3-

month-old male, and Pt. 3.3 was a 5-year 2-month-old female. The remaining children (mean age = 5 years 4.7 months; SD = 1.2 months) included Pt 3.4, Pt. 3.5 and Pt 3.6 and were designated the non-training group (these children were assessed before and after the training phase but received no training). Pt. 3.4 was a 5-year 4-month-old male, Pt. 3.5 was a 5-year 6-month-old male, and Pt. 3.6 was a 5-year 4-month-old female.

Materials

Participants were exposed to a number of different assessments before the training intervention. These included two standardised measures, namely, the Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn, 2007) and the Children’s Category Test (CCT; Boll, 1993); a test of class inclusion (Piaget, 1952); and two tests of relational responding, namely, non-arbitrary containment and arbitrary containment as employed in Study 2. They received all the same assessments after the training intervention except for the test of non-arbitrary containment relations. Three of the participants were exposed to training in arbitrary containment relations during the training phase, which drew on materials used in the arbitrary containment assessment.

PPVT-4. This test assesses receptive vocabulary in respondents aged from 2 years 6 months to late adulthood and is often used as a test of scholastic aptitude. It is administered individually and is presented in a multiple-choice format in which the respondent is presented with four pictures, and asked to select the one that best illustrates the definition of a particular word. The stimuli include items representing up to 20 content areas (e.g., actions, vegetables, tools), and components of speech (nouns, verbs, or attributes), encompassing a broad range of difficulty. Test-retest reliability has been shown to be between .92 and .96

(Community-University Partnership for the Study of Children, Youth, and Families, 2011), while internal consistency has been noted to be similarly high (between .94 and .95). Construct and convergent validity has been assessed by comparing the PPVT-4 to the Expressive Vocabulary Test, Second Edition (EVT-2; Williams, 2007). Correlations between the two were high (r = .80 to .84) for all age groups (Community-University Partnership for the Study of Children, Youth, and Families, 2011).

CCT. This test was designed to assess categorisation skills for children between 5 and 8 years (Level 1) and between 9 and 16 years (Level 2). It is individually administered and is designed to assess concept formation and problem-solving abilities with novel material. In the current study, only Level 1 was used. This level consists of 5 subtests and includes 80 questions. Within each subtest children are asked to determine the principle underlying correct performance (i.e., matching based on colour, size or proportion) using examiner feedback. Previous research has indicated that the test has adequate test-retest reliability (r = .75), in addition to strong internal consistency for both levels 1 (r = .88) and 2 (r = .86; MacNeill, 1996).

Class Inclusion Test. This is a classic test of categorisation ability that assesses the capacity to respond to a stimulus as simultaneously belonging to both a class and a superordinate class (Thomas & Horton, 1997). It has its origins in Piagetian developmental psychology (Piaget, 1952) where it was seen as determining if a child had reached the “concrete operational” stage of development, hypothesised to occur by age seven or eight. In the test, the child is first shown an array of stimuli, which belong to a particular category. These stimuli form two different subclasses, with a greater number of stimuli within one category than the other (e.g., 5 apples and 2 pears). The child is asked if there are more of the larger subclass or of the category; for example "Are there more apples or are there more fruit?" Materials used in the current study included 5.5cm x 5.5cm colored flashcards depicting a variety of animals (horse, dog, pig, cow, cat, and dog), fruit (strawberry, apple, pear, banana, lemon, orange), clothing (socks, skirt, t-shirt, dress, trousers, coat) and vehicles (car, fire engine, truck, motorbike, tractor, bus). Multiple examples of each of these stimuli were employed.

Non-Arbitrary Containment Assessment. The materials used for this assessment were identical to the non-arbitrary containment assessment outlined in Study 2.

Arbitrary Containment Assessment. The materials employed within this assessment and in the training phase were identical to those used in Study 2.

Experimental Design

The current study employed a combined concurrent multiple baseline design (multiple probe design across behaviours, and a multiple baseline design across participants) with three participants (i.e., Pt. 3.1, Pt. 3.2, Pt. 3.3). All participants were exposed to baseline sessions until a stable level of responding was observed (meaning that scores for the last three probes were equal or that the majority of scores among the last five probes were equal). Following this, one participant (i.e., Pt. 3.1) was entered into Phase 1 of training, while the remaining participants remained in baseline conditions. During each phase of training (i.e., mutual entailment, transformation of function via mutual entailment, combinatorial entailment, and transformation of function via combinatorial entailment) the remaining untrained relational components were probed during experimental sessions. The two remaining participants were exposed to baseline sessions until the first participant had completed the first phase of training. Thereafter, the next participant (Pt. 3.2) was exposed to training once they had shown stable rates of responding. After that second participant had completed the first phase of training, and again assuming stable responding, then the third participant (Pt. 3.3) was exposed to training.

In addition to the advantages offered by a multiple baseline design across responses outlined previously, there are a number of advantages associated with a multiple baseline design across participants. For instance, this design is particularly suitable for use within educational and applied research when three or more clients or participants demonstrate similar behavioural repertoires (i.e., either excessive or deficient) that require intervention. Additionally, multiple baseline designs across participants are especially useful when time and resources are limited. For example, introducing a training intervention to one participant at a time is practical as it allows the practitioner to carefully manage the investment of their time (and the participants' time), thereby allowing for the timely termination of an unsuccessful intervention, ensuring that neither the practitioner or participant spends unnecessary time with such an intervention. Finally, educational and applied research aims to identify interventions and educational programs that are effective with a number of individuals. As such, a multiple baseline across participants design is particularly useful when considering this goal (Gast et al., 2014).

Procedure

The school provided the setting for the work and all assessment and testing took place during school hours in a classroom separate from the main one in which children's normal lessons took place. Participants were assessed and trained in individual sessions. All participants were initially assessed using the PPVT-4, CCT, and class inclusion test across three sessions ranging in duration between 15 and 40 minutes each. As with Study 2, all participants were assessed for non-arbitrary and arbitrary containment relations over two sessions, each of which lasted roughly 40 minutes.

Assessing Non-Arbitrary Containment. All participants were assessed for non-arbitrary containment relations as per Study 2.

Assessing Arbitrary Containment. All participants were assessed for arbitrary containment relations using a laptop-based PowerPoint presentation as per Study 2.

Training Arbitrary Containment. The three training participants (Pt. 3.1, Pt. 3.2, Pt. 3.3) were exposed to arbitrary containment relational training over the course of a number of weeks. Each training session lasted approximately 40 minutes and sessions were held five times per week.

As explained in the design section, all three participants were exposed to baseline sessions before receiving training in accordance with the sequential logic of the multiple baseline design. The baseline assessment and arbitrary containment training was identical to the assessment and training of arbitrary containment previously described in Study 2.

Generalisation Assessment. Following the completion of each phase of experimental training, the participant was then exposed to generalisation probes using Stimulus Sets 2, 3 and 4 as per Study 2. Generalisation probes were identical in format to baseline assessments and were also comprised of 64 trials. No feedback was provided in these trials. Criterion for demonstration of generalisation was 90% correct (see Study 2 for rationale regarding generalisation criterion).

Maintenance Assessment. Five weeks following the completion of the fourth phase of experimental training, the participant was exposed to maintenance probes using Stimulus Set 1 (maintenance of training with trained stimulus sets), and Stimulus Sets 2 – 4 (maintenance of training with untrained stimulus sets). As such, any failure to maintain the trained arbitrary containment repertoire could be separated from a failure to maintain the generalised skill. Maintenance probes were identical in format to baseline assessments and were comprised of 64 trials. No feedback was provided in these trials. Criterion for demonstration of maintenance was 90% correct (see Study 2 for rationale regarding maintenance criterion).

Post-Training. All participants (both Training and Non-training) were again assessed for their verbal ability (PPVT-4), classification ability (CCT) and class inclusion responding 6 months after the initial assessments.

Procedural Fidelity and Inter-Observer Agreement (IOA). Sessions were observed in person for IOA by a secondary data collector for 36.7% of all trials. A total agreement score of 99.77% was reached (96.88% - 100%). Procedural fidelity data were collected for 36.7% of all trials. The results indicated that procedural integrity ranged from 96.88% - 100% (average 99.42%).

Results

The initial relational assessment showed that all six participants (i.e., both Training and Non-training) met criterion (i.e., achieved over 80% correct) for the non-arbitrary containment assessment overall (see Table 4.6). At the same time, all six failed the arbitrary containment assessment both overall (with no participant scoring over 60%) as well as at the level of each of the relational components (see Table 4.7).

The results of the combined multiple baseline design are displayed in Figure 4.2. Baseline scores for the first relational component, mutual entailment, remained stable for all three participants prior to the introduction of the initial phase of arbitrary containment training. The sequential introduction of training (both for mutual entailment as well as for other components) for each participant resulted in increases to criterion for that participant. In the case of each participant, the sequential introduction of training for each relational component resulted in increases to criterion for that component. These increases were obtained without corresponding changes in the baseline of the remaining untrained relational components and/or participants. The results of the multiple baseline analysis for generalisation across untrained stimuli and maintenance effects for both trained and untrained stimuli are also presented in Figure 4.2. These data show universally successful generalisation and maintenance of the relational repertoires.

Table 4.6. Percentage of correct non-arbitrary containment relational responses.

Participant	Mutual Entailment	ToF Mutual Entailment	Combinatorial Entailment	ToF Combinatorial Entailment	Total Score
Training Participants					
Pt. 3.1	100%	100%	100%	100%	100%
Pt. 3.2	100%	100%	100%	93.75%	98.43%
Pt. 3.3	100%	100%	100%	87.5%	96.87%
Comparison Participants					
Pt. 3.4	100%	100%	100%	100%	100%
Pt. 3.5	100%	100%	100%	93.75%	98.43%
Pt. 3.6	100%	100%	93.75%	81.25%	93.75%

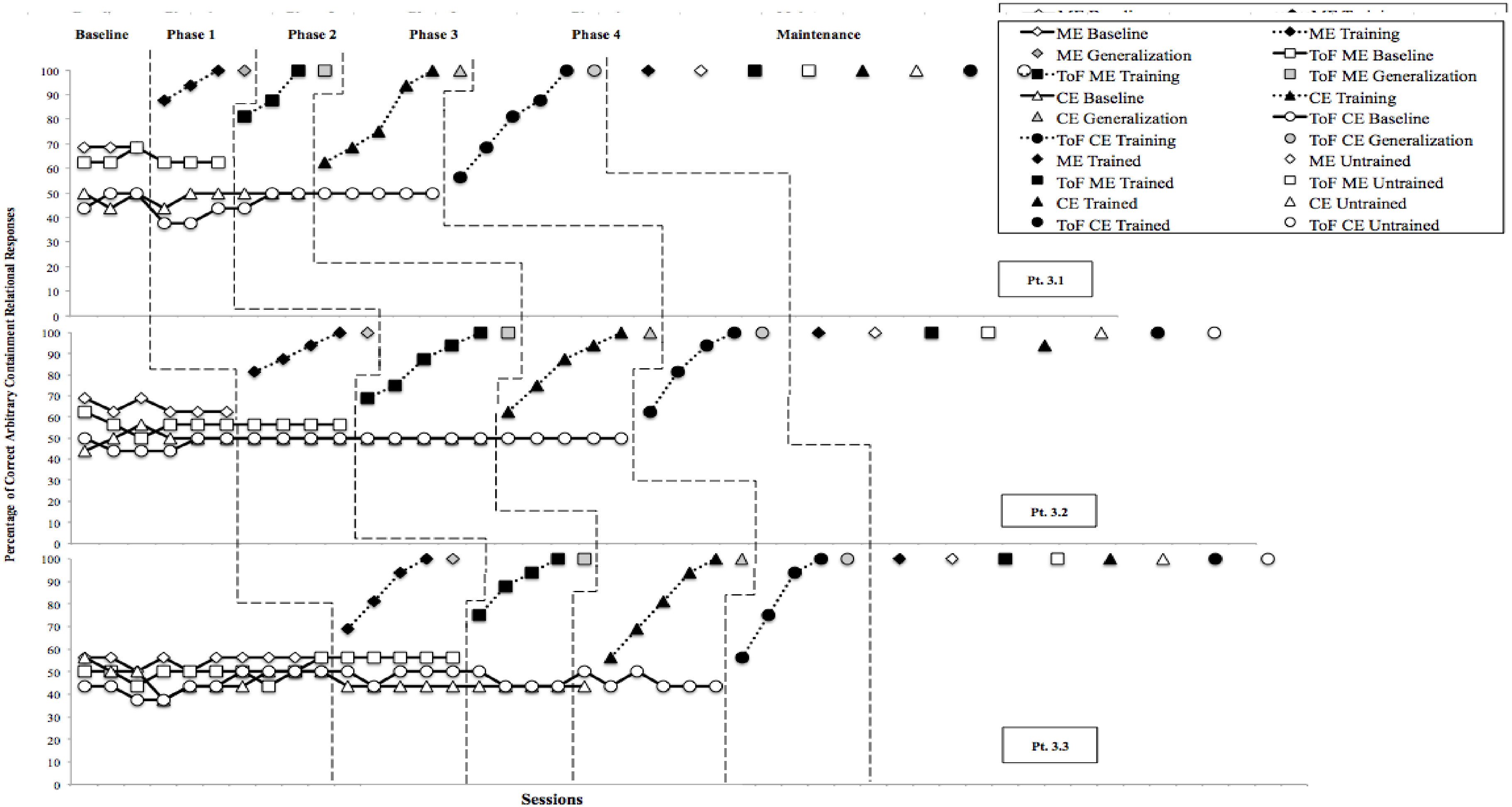
Note. ToF = Transformation of Stimulus Function.

Table 4.7. Percentage of correct arbitrary containment relational responses.

Participant	Mutual Entailment	ToF Mutual Entailment	Combinatorial Entailment	ToF Combinatorial Entailment	Total Score
Training Participants					
Pt. 3.1	68.75%	62.5%	50%	43.75%	56.25%
Pt. 3.2	62.5%	62.5%	50%	43.75%	54.69%
Pt. 3.3	56.25%	50%	50%	43.75%	50%
Comparison Participants					
Pt. 3.4	75%	56.25%	50%	50%	59.38%
Pt. 3.5	62.5%	56.25%	50%	50%	57.81%
Pt. 3.6	62.5%	56.25%	43.75%	37.5%	50%

Note. ToF = Transformation of stimulus function.

Figure 4.2. Combined multiple baseline across components and across participants design displaying correct arbitrary containment responses.



Note. ME = Mutual entailment; ToF ME = Transformation of function of mutually entailed relations; CE = Combinatorial entailment; ToF CE = Transformation of function of combinatorially entailed relations.

Table 4.8. Pre-Training and Six-Month Post-Training Raw scores of Training and Comparison Participants.

Participant	**PPVT Raw Score**			**CCT Raw Score**			**CI Score**		
	T1	**T2**	**Score Change**	**T1**	**T2**	**Score Change**	**T1**	**T2**	**Score Change**
Training Participants									
Pt. 3.1	87	95	+8	66	69	+3	4	5	+1
Pt. 3.2	84	90	+6	62	64	+2	4	5	+1
Pt. 3.3	81	89	+8	60	64	+4	3	5	+2
Comparison Participants									
Pt. 3.4	88	90	+2	66	67	+1	5	5	0
Pt. 3.5	89	92	+3	67	67	0	5	5	0
Pt. 3.6	85	87	+2	66	67	+1	3	4	+1

Note. PPVT-4 Raw Score = Peabody Picture Vocabulary Test, 4th Edition raw score; CCT Raw Score = Children's Category Test raw score; CI = Class Inclusion; Time 1 = Pre-training; Time 2 = 6 months post-training.

Table 4.9. Correct arbitrary containment scores at intake and six months post-training.

Participants	ME		ToF ME		CE		ToF CE		Total Score	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
Training Participants										
Pt. 3.1	68.75%	100%	62.5%	100%	50%	100%	43.75%	100%	56.25%	100%
Pt. 3.2	62.5%	100%	62.5%	100%	50%	100%	43.75%	93.75%	54.69%	98.44%
Pt. 3.3	56.25%	100%	50%	100%	50%	100%	43.75%	100%	50%	100%
Comparison Participants										
Pt. 3.4	75%	81.25%	56.25%	56.25%	56.25%	56.25%	50%	50%	59.38%	60.94%
Pt. 3.5	62.5%	68.75%	56.25%	62.5%	50%	56.25%	50%	50%	57.81%	59.38%
Pt. 3.6	62.5%	68.75%	56.25%	56.25%	43.75%	50%	37.5%	37.5%	50%	53.13%

Note. ME = Mutual entailment; ToF = Transformation of function; CE = Combinatorial entailment; T1 = Time 1 (intake score); T2 = Time 2 (six months-post training).

As an example of individual data, we can consider the performance of Pt. 3.1. The baseline mutual entailment scores for this participant remained at 11/16 (68.75%) throughout baseline testing. Following the introduction of arbitrary containment training for mutual entailment (Phase 1), this participant's performance increased substantially and met criterion within three sessions. They then showed successful generalisation to untrained stimuli. Similarly successful outcomes were seen in the case of each of the other relational components (i.e., transformation of function via mutual entailment, combinatorial entailment, and transformation of function via combinatorial entailment). This participant also passed maintenance testing, conducted some 5 weeks after the end of the intervention, for each of the relational components and with both trained and untrained stimuli in each case. Similarly successful patterns to those shown by Pt. 3.1 were seen in the case of each of the other two participants (i.e., Pt. 3.2 and Pt. 3.3) also.

Both Training and Non-training group participants were assessed for PPVT-4, CCT and class inclusion scores prior to training, as well as six months following the cessation of training. The raw scores for these measures for each of the two time-points are shown in Table 4.8. The results of this assessment indicated that while both the Training and Non-training groups evidenced gains in responding in each of the domains in question, the gains made by the Training group were greater than those of the Non-training group across each of the three instruments. More specifically, the Training and Non-training groups showed mean improvements of 7.3 and 2.3 respectively in the PPVT-4, 3 and 0.67 respectively in the CCT And 1.3 and 0.3 respectively in the class inclusion test.

Both Training and Non-training participants were tested for arbitrary containment repertoires six months following the cessation of training. The results indicated that all Training participants maintained high levels of responding while the Non-training participants showed minimal gains in performance when compared to testing at intake (see Table 4.9).

Discussion

Study 3 extends and replicates the findings of Study 2 with additional evidence regarding the impact of the relational frame training involved in both. The research contained within the current chapter also extends the findings of previous research (e.g., Barnes-Holmes, Barnes-Holmes, Smeets & Luciano, 2004; Barnes-Holmes, Barnes-Holmes, Smeets, Strand & Friman, 2004; Berens & Hayes, 2007) by providing additional controlled evidence that relational frames are learned and can be considered to be generalised operants which can be established via a process of MET and contingent feedback.

Studies 2 and 3 provided evidence that containment relational framing, which may also contribute to classification responding, may be trained with typically developing young children. In Study 2, one 5-year-old girl was successfully trained in arbitrary containment relational framing using a multiple baseline design across components of the relational framing repertoire, while three children in the 5-year-old age range were successfully trained in arbitrary containment relational framing using a multiple baseline design across both components in the relational framing repertoire as well as participants in Study 3. These data contribute to previous empirical RFT research indicating that repertoires of relational framing can be trained via MET (e.g., Barnes-Holmes, Barnes-Holmes & Smeets, 2004; Barnes-Holmes, Barnes-Holmes, Smeets, Strand & Friman, 2004; Berens & Hayes, 2007). In particular, Study 3 also contributes to earlier research (i.e., Berens & Hayes, 2007) that has accomplished this within the context of a well-controlled experimental design (i.e., multiple baseline design across components and participants). As such, Study 3 has provided additional evidence regarding the operant nature of relational framing and its components (i.e., mutual entailment, combinatorial entailment and transformation of stimulus functions). The data from these studies constitute a potentially important extension of previous RFT work by demonstrating the acquisition, generalisation and maintenance of the containment framing repertoire. The examination of maintenance of containment responding is particularly important as maintenance has not typically been assessed for relational framing repertoires despite being an important goal of training protocols. Both Studies 2 and 3 showed a high level of maintenance of up to five weeks for containment responding for both trained (Study 2 and 3) and untrained (Study 3) stimulus sets. This outcome further indicates the efficacy of MET when establishing relational framing repertoires.

In addition to this, the results of Study 3 indicate that training arbitrary containment framing can perhaps augment intellectual performance. More specifically, in Study 3, the children who were exposed to arbitrary containment training demonstrated improvements in their performance on tests of linguistic and cognitive ability. The pattern of improved performance across these assessments was markedly better for children exposed to training that that seen in the case of a control group of children who had not received training. This result is in line with previous RFT studies which have demonstrated that relational frame training can improve performance on standardised cognitive assessments (see for example, Cassidy, Roche & Hayes, 2011; Stewart & Hayes, 2016). Although the current data are generated from a small sample size, these findings are promising and also indicate the need for future research in this avenue.

In addition, Study 3 indicated an improvement in performances on tests of classification. These tests were administered because previous research (i.e., Study 2 contained in Chapter 3) indicated that the framing repertoire being trained in this case might be relevant to classification. At the same time, it is possible that training any relational frame might positively impact performance on such assessments. Thus, future research is necessary to determine whether specific frames may be more likely than others to boost performance within specific intellectual domains (e.g., classification as measured by the CCT) and to also determine whether training containment may support classification responding in addition to other behavioural repertoires.

The findings generated from Study 3 potentially have significant connotations for applied behaviour analysis and education. For instance, this study indicates a specific relational frame to train that may be central to both language and cognition. Identifying the relational frame that is fundamental to specific behavioural repertoires (e.g., coordinate relational framing may be fundamental to matching) is important when considering a program to establish these repertoires. Both Studies 2 and 3 indicate the efficacy of an RFT-based training procedure to facilitate derived relational responding repertoires in accordance with containment among young children. Furthermore, the results generated from Study 3 provides additional support for the relationship between derived relational responding and language and cognition (Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; Mulhern, Stewart & McElwee, 2017).

Additionally, Study 3 presents a potentially useful training protocol to establish containment relational responding in children or other populations.

The success of these studies also indicates that the training protocol offers a possible framework for additional research assessing and establishing relational framing of containment and classification. Such research could address the limitations of these studies in addition to expanding on previous research in this domain. For instance, one limitation which future research should address is to identify the extent to which such training can impact cognitive performance. The data from Study 3 fail to provide definitive conclusions in this respect due to the very small sample sizes employed and as such do not provide conclusive evidence regarding the impact of arbitrary containment training on general cognitive performance or more specifically, the performance of categorisation. The small sample size used in Studies 2 and 3 was appropriate for the combined multiple baseline design that was employed to demonstrate that MET could facilitate specific patterns of relational framing consistent with containment with 5- to 6-year-old children. However, in the case of Study 3, the sample size employed was too small to provide accurate conclusions regarding comparisons between the training and non-training groups. Future research could advance research in this field by delivering this training program in a larger between-group design.

In conclusion, the findings of Studies 2 and 3 indicate that relational framing of containment can be successfully trained in young children in whom such repertoires are absent or deficient. The results of Study 3 also indicate that arbitrary containment training can improve performances on standardised measures of cognition and language. However, additional research is required in order to determine the degree to which such training impacts cognitive performance in general as well as categorisation repertoires.

Chapter 5

Studies 4 and 5

Facilitating Arbitrary Hierarchy Repertoires in Young Children*

* Portions of this chapter have been submitted for publication:

Mulhern, T., Stewart, I., & McElwee, J. (under submission). Facilitating relational framing of classification in young children.

According to Barnes-Holmes and colleagues (2004) relational framing skills form the foundation of a spectrum of cognitive abilities correlated with educational achievement including classification. From this point of view, classification can be conceptualised as involving particular types of framing (e.g., containment, hierarchy, etc.). Chapter 4 was concerned with examining and facilitating the frame of containment, which is theorised to develop prior to more complex hierarchical framing (Hayes et al., 2001), while the current chapter focuses on arbitrary hierarchy alone. As with non-arbitrary containment and arbitrary containment, arbitrary hierarchical responding initially develops with mutual non-arbitrary hierarchical relations. For example, a child could learn to label one object as being a member of a particular stimulus set (e.g., “a terrier is a type of dog”) and to also label this object as being contained within a class of objects (e.g., “the class of dogs contains terriers”). This repertoire might then come under contextual control (i.e., of cues including the words “type of”, “includes”, “member of”, and “contains”) and then generalise. The combination of these frames (i.e., non-arbitrary containment, arbitrary containment, and arbitrary hierarchy) may form the basis for relatively more abstract patterns of responding, such as classification (i.e., responding to “members” as being contained in “classes”) and hierarchical classification (classes being contained within classes).

Gil, Luciano, Ruiz and Valdivia-Salas (2012) were the first to conduct a relational frame exploration of hierarchy. Within the first phase of training arbitrary stimuli were established as contextual cues for hierarchical relational responding (i.e., “includes”, and “belongs to”) by teaching participants to perform non-arbitrary containment relational responding when presented with these stimuli. Following this, the experimenters then employed a similar procedure to establish contextual cues for responding consistent with “same” and “different” relations by teaching participants to select physically identical stimuli when presented with one arbitrary shape, and to also select physically different stimuli when presented with another arbitrary stimuli. The experimenters then taught and assessed all four cues (i.e., “includes”. “belongs to”, “same”, and “different”) in tasks using stimuli in verbal relations which the experimenters assumed to have been previously established with the participants (e.g., when provided with the contextual cue for “includes” the participant was reinforced for selecting the word “vowels” when presented with the word “alphabet”). Within the second experimental phase, the participants were then taught three 4-member equivalence relations involving abstract shapes and nonsense syllables by employing a standard one-to-many matching-to-sample (MTS) procedure. Participants were subsequently tested for these equivalence relations. In Phase 3, the experimenters employed the previously established contextual cues of “includes” and “belongs to” to form what was assumed to be an arbitrary hierarchical relational network. This network was composed of some of the stimuli within the equivalence classes established during the previous phase of training, which comprised the bottom level of the hierarchy, while the middle and upper levels of the hierarchy were comprised of novel abstract shapes. In Phase 4, participants were then taught associations between specific stimuli in the arbitrary relational network previously formed in Phase 3, in addition to words describing specific stimulus properties (e.g., “sweet”, and “cold”) using a stimulus pairing and MTS procedure. The final experimental phase (Phase 5) involved testing participants for predicted transformation of stimulus functions consistent with hierarchical relations of stimuli that had not been explicitly connected with specific properties within the previous experimental phase.

The results of this study indicated that nine of the ten undergraduate participants demonstrated the expected pattern of transformation of stimulus functions in Phase 5. Gil and colleagues (2012) is an important investigation of hierarchical relational framing, particularly Phase 1, which demonstrates several empirical and methodological procedures (e.g., establishing contextual cues for containment relations) and also demonstrated a format in which responding consistent with several stimulus relations is assessed by requiring participants to select contextual cues for specific frames. However, Gil and colleagues' demonstration of hierarchical relational framing also has several limitations. Firstly, Gil and colleagues failed to assess any potential derived relations in the arbitrary hierarchical relational network formed within the third phase of testing. Given the fact that relational frames are defined by patterns of derived relations, such as mutual entailment and combinatorial entailment, the failure to assess these properties calls into question the demonstration of hierarchical relational framing within the study. Secondly, the study lacks some experimental control, as a number of the key contexts for teaching and assessing hierarchical relational responding included stimuli in extra-experimentally formed hierarchical relations with one another. This constitutes a significant problem in regards to the empirical exploration of any form of relational framing, and is particularly relevant within applied work, which may seek to establish arbitrary hierarchical relational repertoires among young children, or populations with developmental disabilities.

Gil and colleagues (2014) furthered their contribution to the understanding of hierarchical relational framing by analysing the transformation of stimulus function via hierarchical relations with eight adults using a computer task comprised of five phases. The stimuli employed were the same as those within Gil and colleagues' original experiment (2012). Arbitrary symbols were employed as the relational cues "includes", "belongs to", "same", and "different". During the first experimental phase, four of these stimuli were employed and trained to serve as the aforementioned relational cues. Within the second experimental phase, three equivalence classes were taught and assessed (A1-B1-C1-D1; A2-B2-C2-D2; A3-B3-C3-D3). Following this, inclusion relations were then established within Phase 3, by employing the "includes" and "belongs to" relational cues between the lower levels of the hierarchy (i.e., A1/B1, A2/B2, and A3/B3) and stimuli X.1, X.2 and Y.1 respectively. The "includes" relational cue was then employed to establish inclusion relations between X.1/X.2 and X, and between Y.1 and Y, so that X and Y served as the most inclusive levels of two separate hierarchical networks. The experimenters then established X.1 as cold, D2 as heavy, and C3 as sweet within the fourth experimental phase. Finally, participants were then assessed for the transformation of stimulus function using seven stimuli from both hierarchical networks within Phase 5. Of the six participants who reached this phase of testing, five responded correctly. The findings of Gil and colleagues (2014) experiment presented several advantages over the previous 2012 experiment. Firstly, in Gil and colleagues (2012) experiment, derived relations were not distinctly isolated as the hierarchical relations were taught in both directions (e.g., X includes X.2, and X.2 belongs to X). However, Gil and colleagues (2014) rectified this by establishing the hierarchical relational networks in only one direction (e.g., X includes X.2), thereby providing a more robust demonstration of the transformation of stimulus functions in accordance with hierarchical relations. Secondly, the response options in the 2014 experiment isolate a more exact transformation of incidental or non-inclusive features of the stimuli contained within hierarchy X. The researchers achieved this by omitting the word "always" from the final transformation of stimulus function test, which was present in the initial 2012 experiments. Furthermore, the experimenters added the cue "part of" to the assessment of transformation of function, which served to isolate the transformation of stimulus functions from the lower to the higher level of the hierarchy. However, the study also has a number of limitations. Firstly, the participants transformation of stimulus function responses were assessed using seven stimuli across seven trials and the researchers themselves outline that the test structure may have influenced correct responding. As such, future studies should endeavour to include a greater number of trials to test for this property. Finally, the researchers outline that the stimuli employed within the experiment may have obscured the isolation of the relevant functions to establish the hierarchical cues, as the type of trials and stimuli employed throughout the experimental phases were primarily based on pre-experimental functions (e.g., the picture of a face and the pictures of a nose, a mouth, and an eye) in order to facilitate the acquisition of the intended inclusion/belonging function for these arbitrary cues.

Slattery and Stewart (2014) sought to address the limitations of Gil et al.'s study (2012) by addressing hierarchical relational framing to model hierarchical classification as contextually controlled generalised relational responding. The first experiment involved non-arbitrary relational training via conditional discriminative training in which contextual control was established using stimuli, which shared physical features with one another. By doing so, the experimenters established two arbitrary shapes as contextual cues for “member of” and “includes” relational responding. Within this phase of training, participants were taught to select specific shapes with particular features when presented with particular nonsense stimuli (e.g., H1 → circular stimuli; H2 → square stimuli; H1.1 → a circular blue stimulus; H1.2 → a circular yellow stimulus; H2.1 → a square blue stimulus; H2.2 → a square yellow stimulus). Participants were taught to relate specific classes of shapes with specific nonsense syllables. Participants were also taught to relate nonsense syllables in specific ways when presented with arbitrary shapes with served as contextual cues (i.e., “member of”, “includes”). For example, when presented with the arbitrary shape which served as the contextual cue “member of” and with H1.1 in an array of stimuli, the participant was reinforced for selecting H1. Similarly, when presented with the arbitrary shape serving as the contextual cue “includes” and H1 within an array of stimuli the participant was reinforced for selecting H1.1. These cues were subsequently employed to teach a hierarchical relational network using a novel set of arbitrary stimuli. The researchers then assessed participants for the derivation of additional untrained hierarchical relations (i.e., mutual entailment and combinatorial entailment) in addition to transformation of stimulus function. The resultant patterns of relational framing demonstrated mutual entailment, combinatorial entailment, and transformation of stimulus function among the adult participants. Experiment 2 was similar to that of Experiment 1, however, it extended the basic model via the addition of more naturalistic stimuli and also adjusted the assessment for transformation of stimulus function among adult participants. The results from this experiment indicate that several further participants were successfully taught and assessed for both non-arbitrary and arbitrarily applicable hierarchical relational responding. However, the results of Experiment 2 differed from those of Experiment 1 as there were more failures in the non-arbitrary phase in Experiment 2, while there were no failures in Experiment 1. Furthermore, the expected pattern of uni-directional transformation of stimulus functions was not demonstrated as strongly across those participants who reached the arbitrarily applicable relational phase. While this study represents a significant step forward in the assessment and training of hierarchical relational responding, the results are limited as the participants employed were university students who presumably already had established hierarchical framing repertoires. As such, the current study aims to extend the research in the field of RFT by assessing and training hierarchical framing repertoires among typically developing children. Study 4 is a pilot study which aims to augment arbitrary hierarchical relational repertoires with a 6-year-old child in the context of a multiple baseline across responses design. Study 5 provides a further examination of an MET approach to facilitate arbitrarily applicable hierarchical responses within a combined multiple baseline design across responses and a multiple baseline design across participants similar to that of Berens and Hayes (2007), and further determines the impact of arbitrary hierarchical training on additional outcome measures, including language, categorisation, and class inclusion.

Study 4 Method

Participant & Setting.

Prior ethical approval for the recruitment of participants for this study was obtained from the research ethics committee of the National University of Ireland, Galway. A typically developing 6-year 4-month-old Irish female participated in the current pilot study (Pt. 4.1). Parental consent was obtained for the child and verbal consent was also obtained from the participant herself.

The participant was selected base on her performance on an assessment of arbitrary hierarchical relational responding (see Materials) in which she failed by achieving a score of 60% or less. The score of 60% or less of correct arbitrary hierarchy responses was selected as the results of Study 1 indicated that this rate of responding (or lower) was indicative of a weak and unestablished repertoire that was no greater than chance responding.

Baseline, training, generalisation and maintenance sessions were conducted within the participants' own home within a quiet room. Sessions occurred between three and five times per week for 40-minute sessions.

Materials

The participant was assessed for arbitrary hierarchy before being exposed to training in arbitrary hierarchical relations during the training phase, which also drew on materials used in the arbitrary hierarchy assessment.

Arbitrary Hierarchy Assessment. The materials included four stimulus sets of nonsense syllables (Table 5.1), which were presented as textual stimuli on a laptop screen using PowerPoint. The participant was assessed using all four stimulus sets (i.e., Sets 1 – 4) and was trained using Stimulus Set 1 alone.

Table 5.1. The table represents the different stimulus sets used in arbitrary hierarchical training.

	Stimulus Set 1	**Stimulus Set 2**	**Stimulus Set 3**	**Stimulus Set 4**
A	Zimp	Ving	Sibe	Fouk
B	Yalt	Unda	Ropa	Hib
C	Wilk	Taga	Quilb	Klab

Experimental Design

The current study employed a multiple baseline design across responses (for the benefits of this experimental design, see Chapter 4). The first of these responses, mutual entailment of arbitrary hierarchical relational responding, was assessed until a stable level of responding was observed (meaning that scores for the last three probes were equal or that the majority of scores among the last five probes were equal). Following this Pt. 4.1 was entered into Phase 1 of training, while the remaining responses (i.e., transformation of stimulus function via mutual entailment, combinatorial entailment, and transformation of stimulus function via combinatorial entailment) were probed within baseline sessions. This sequence of testing was repeated for (a) transformation of stimulus function via mutual entailment, (b) combinatorial entailment, and (c) transformation of stimulus function via combinatorial entailment in which the participant was exposed to five, ten, and thirteen baseline sessions respectively.

Procedure

The participant was assessed and trained in individual sessions within her own home. She was assessed for arbitrary hierarchy relations within one session, which lasted roughly 40 minutes, in order to determine eligibility for the study.

Assessing Arbitrary Hierarchy. The participant was assessed for arbitrary hierarchical relations using a laptop-based PowerPoint presentation involving nonsense syllables (see Table 5.2 for arbitrary hierarchy trial types). The participant was assessed for (a) mutual entailment, (b) transformation of function via mutual entailment, (c) combinatorial entailment, and (d) transformation of function via combinatorial entailment.

Table 5.2. Trial types used to assess and train arbitrary hierarchy.

Property	Statement	Question [Answer]
Mutual Entailment (2 stimuli)	[A] is a type of [B]	Is [A] a type of [B]? [Yes]
	[A] is a type of [B]	Is [B] a type of [A]? [No]
	[A] is a type of [B]	Does the class [A] contain [B]? [No]
	[A] is a type of [B]	Does the class [B] contain [A]? [Yes]
	The class [A] contains [B]	Is [A] a type of [B]? [No]
	The class [A] contains [B]	Is [B] a type of [A]? [Yes]
	The class [A] contains [B]	Does the class [A] contain [B]? [Yes]
	The class [A] contains [B]	Does the class [B] contain [A]? [No]
	[B] is a type of [A]	Is [A] a type of [B]? [No]
	[B] is a type of [A]	Is [B] a type of [A]? [Yes]
	[B] is a type of [A]	Does the class [A] contain [B]? [Yes]
	[B] is a type of [A]	Does the class [B] contain [A]? [No]
	The class [B] contains [A]	Is [A] a type of [B]? [Yes]
	The class [B] contains [A]	Is [B] a type of [A]? [No]
	The class [B] contains [A]	Does the class [A] contain [B]? [No]
	The class [B] contains [A]	Does the class [B] contain [A]? [Yes]
Transformation of Function (2 stimuli)	[A] is a type of [B]. [A]'s have [Property A], [B]'s have [Property B].	Do all [A]'s have [Property B]? [Yes]
	[A] is a type of [B]. [A]'s have [Property A], [B]'s have [Property B].	Do some [B]'s have [Property A]? [Yes]
	[A] is a type of [B]. [A]'s have [Property A], [B]'s have [Property B].	Does the class A contain members without [Property B]? [No]
	The class [A] contains [B]. [A]'s have [Property A], [B]'s have [Property B]	Do all [B]'s have [Property A]? [Yes]
	The class [A] contains [B]. [A]'s have [Property A], [B]'s have [Property B]	Do some [A]'s have [Property B]? [Yes]
	The class [A] contains [B]. [A]'s have [Property A], [B]'s have [Property B]	Does the class [B] contain members without [Property A]? [No]
	[B] is a type of [A]. [A]'s have [Property A], [B]'s have [Property B]	Do all [B]'s have [Property A]? [Yes]
	[B] is a type of [A]. [A]'s have [Property A], [B]'s have [Property B]	Do some [A]'s have [Property B]? [Yes]
	[B] is a type of [A]. [A]'s have [Property A], [B]'s have [Property B]	Does the class [B] contain members without [Property A]? [No]
	The class [B] contains [A]. [A]'s have [Property A], [B]'s have [Property B]	Do all [A]'s have [Property B]? [Yes]
	The class [B] contains [A]. [A]'s have [Property A], [B]'s have [Property B]	Do some [B]'s have [Property A]? [Yes]
	The class [B] contains [A]. [A]'s have [Property A], [B]'s have [Property B]	Does the class A contain members without [Property B]? [No]
Combinatorial Entailment (3 stimuli)	[A] is a type of [B], [B] is a type of [C]	Is [A] a type of [C]? [Yes]
	[A] is a type of [B], [B] is a type of [C]	Is [C] a type of [A]? [No]
	[A] is a type of [B], [B] is a type of [C]	Does the class [A] contain [C]? [No]
	[A] is a type of [B], [B] is a type of [C]	Does the class [C] contain [A]? [Yes]
	The class [A] contains [B], the class [B] contains [C]	Is [A] a type of [C]? [No]
	The class [A] contains [B], the class [B] contains [C]	Is [C] a type of [A]? [Yes]
	The class [A] contains [B], the class [B] contains [C]	Does the class [A] contain [C]? [Yes]
	The class [A] contains [B], the class [B] contains [C]	Does the class [C] contain [A]? [No]
	[C] is a type of [B], [B] is a type of [A]	Is [A] a type of [C]? [No]
	[C] is a type of [B], [B] is a type of [A]	Is [C] a type of [A]? [Yes]
	[C] is a type of [B], [B] is a type of [A]	Does the class [A] contain [C]? [Yes]

	[C] is a type of [B], [B] is a type of [A]	Does the class [C] contain [A]? [No]
	The class [C] contains [B], the class [B] contains [A]	Is [A] a type of [C]? [Yes]
	The class [C] contains [B], the class [B] contains [A]	Is [C] a type of [A]? [No]
	The class [C] contains [B], the class [B] contains [A]	Does the class [A] contain [C]? [No]
	The class [C] contains [B], the class [B] contains [A]	Does the class [C] contain [A]? [Yes]
Transformation of Function (3 stimuli)	[A] is a type of [B], [B] is a type of [C]. [A]'s have [Property A], [C]'s have [Property C]	Do all [A]'s have [Property C]? – [Yes]
	[A] is a type of [B], [B] is a type of [C]. [A]'s have [Property A], [C]'s have [Property C]	Do some [C]'s have [Property A]? [Yes]
	[A] is a type of [B], [B] is a type of [C]. [A]'s have [Property A], [C]'s have [Property C]	Does the class [A] contain members without [Property C]? [No]
	The class [A] contains [B], the class [B] contains [C]. [A]'s have [Property A], [C]'s have [Property C]	Do all [C]'s have [Property A]? [Yes]
	The class [A] contains [B], the class [B] contains [C]. [A]'s have [Property A], [C]'s have [Property C]	Do some [A]'s have [Property C]? [Yes]
	The class [A] contains [B], the class [B] contains [C]. [A]'s have [Property A], [C]'s have [Property C]	Does the class [C] contain members without [Property A]? [No
	[C] is a type of [B], [B] is a type of [A]. [A]'s have [Property A], [C]'s have [Property C]	Do all [C]'s have [Property A]? [Yes]
	[C] is a type of [B], [B] is a type of [A]. [A]'s have [Property A], [C]'s have [Property C]	Do some [A]'s have [Property C]? [Yes]
	[C] is a type of [B], [B] is a type of [A]. [A]'s have [Property A], [C]'s have [Property C]	Does the class [C] contain members without [Property A]? [No
	The class [C] contains [B], the class [B] contains [A]. [A]'s have [Property A], [C]'s have [Property C]	Do all [A]'s have [Property C]? – [Yes]
	The class [C] contains [B], the class [B] contains [A]. [A]'s have [Property A], [C]'s have [Property C]	Do some [C]'s have [Property A]? [Yes]
	The class [C] contains [B], the class [B] contains [A]. [A]'s have [Property A], [C]'s have [Property C]	Does the class [A] contain members without [Property C]? [No]

Note. The codes [A], [B] and [C] signify different nonsense syllables used within each trial. The actual nonsense syllables (e.g., zimp, yalt, wilk, ving, unda, taga etc.) varied quasi-randomly across trials.

On each trial, the participant was first presented with text on a laptop screen that described the arbitrary hierarchical relation between two nonsense syllables; for example "The Zimp is a type of Yalt" (see Table 5.1 for the sets of stimuli used in the relations in this phase). The experimenter read the text aloud to the participant and then asked her a question based on the stimuli involved which could be answered in the form of a "yes" or "no" response (e.g., "Does the class Yalt contain Zimps?"). No feedback was provided regarding performance and no reinforcement was provided for correct or incorrect responding.

In trials for mutual entailment, the relationship between two stimuli was outlined to the participant and she was required to answer a subsequent question with either “yes” or “no” based on the presented relation. A total of sixteen trials (see Table 5.2) was presented in random order. Trials for the transformation of stimulus function via mutual entailment were similar except for the addition of information regarding stimulus function (see Table 5.2). These questions focused on the potential change in stimulus functions of one of the two stimuli in the mutually entailed relation by virtue of being in that relationship. This section was assessed across 12 questions presented in random order.

In trials for combinatorial entailment, the relationship between three stimuli was outlined and the participant was required to answer a subsequent question with either “yes” or “no” based on the presented relation. A total of sixteen trials (see Table 5.2) was presented in random order. Trials for the transformation of function via combinatorial entailment were similar except for the addition of information regarding stimulus function (See Table 5.2). These questions focused on the potential change in stimulus functions of one of the two stimuli in the combinatorially entailed relation by virtue of being in that relationship. This section was assessed across 12 questions presented in random order. The criterion for passing this assessment was showing at least 80% (i.e., approximately 45 out of 56) correct.

Training Arbitrary Hierarchy. Pt. 4.1 was exposed to arbitrary hierarchical relational training over a period of three to four weeks. Each training session lasted approximately 40 minutes and sessions were held five timers per week. The participant was exposed to baseline sessions before receiving training in accordance with the sequential logic of the multiple baseline design. The baseline assessment was identical to the assessment of arbitrary hierarchy relations already outlined.

Training was presented in four phases, corresponding to each of the four components of the relational operant outlined previously in the assessment phase (i.e., mutual entailment, transformation of function via mutual entailment, combinatorial entailment, and transformation of function via combinatorial entailment). Within each phase the remaining untrained relational components were probed (using the same fixed number of trials without feedback of reinforcement as had been used in the assessment phase) during experimental sessions. Training in each phase was the same as assessment of that phase except for the provision of reinforcement and feedback as appropriate and except that trials continued to be presented within a phase until pass criteria for that phase were met (See Appendix C for full training protocol and assessment).

During training sessions, correct responses were reinforced on an FR1 schedule of specific praise relating to the task (e.g., "That's right! A is a type of B"), in addition to an FR1 schedule of tokens (laminated and velcroed pictures which could then be affixed onto a token board), which were then exchanged on an FR4 schedule for a self-chosen sticker. Following incorrect responses the experimenter would say "No, that's not it", and provide contingent feedback (e.g., "B *isn't* a type of A, because A is a type of B!") and the trial would then be represented to the participant until she emitted a correct response. Apart from this, the participant was also told that meeting the goals of each session would result in an additional prize (e.g., rubbers, pencils, markers which she could choose from a box). The goal for the initial training session on all training phases was 50% or more correct. The subsequent goal for each training session was for the participant to beat her previous score by at least one more correct response than the previous training session. The participant was told before each training session how many correct responses she needed to "beat" her previous score and receive the additional prize at the end of the session. She was also informed that if she did not meet this goal, she would not earn a prize for that session. This contingency remained in place until the participant met mastery criterion (i.e., 100% correct in one session, or 90% correct across two consecutive sessions), which signalled the start of the next phase of training.

Generalisation Assessment. Following the completion of each phase of arbitrary hierarchy training, the participant was exposed to generalisation probes using Stimulus Sets 2, 3 and 4. For instance, when Pt. 4.1 met the mastery criterion for Phase 1 she was then assessed for generalised mutually entailed arbitrary hierarchical relational using the untrained stimulus sets. Generalisation probes were identical in format to baseline assessments and were also comprised of 56 trials. No feedback was provided in these trials. Criterion for the demonstration of generalisation was 90% correct. The rate of 90% was selected as the research generated from Study 1 indicated that a strong arbitrary hierarchy repertoire was present if an individual exhibited 80% or more correct arbitrary containment responses. As such, the rate of 90% was selected in order to ensure that the participant had acquired a strong generalized arbitrary hierarchy repertoire.

Maintenance Assessment. The participant was assessed for maintenance of responding for each relational component at one week, two weeks, three weeks, and one month following the mastery of these training phases using Stimulus Set 1. Maintenance probes were identical to that of baseline assessments and were comprised of 56 trials. No feedback was provided in these trials. Criterion for demonstration of maintenance was 90% correct. As outlined previously, the rate of 90% was selected based upon the results generated in Study 1 (see above for rationale).

Inter-Observer Agreement. For 41.9% of all baseline trials, 47.1% of all training trials, 75% of all generalisation trials, and 68.75% of all maintenance trials, a secondary data collector independently scored each trial as correct or incorrect. The secondary data collector was required to reach 100% accuracy on three consecutive mock baseline sessions prior to scoring an experimental session. Agreement data was then collected across all phase types. An agreement was scored when both observers scored a trial as being either correct or incorrect. A disagreement was scored if the observers recorded the trial differently. A percentage agreement score was calculated by dividing the total agreements by the total agreements plus total disagreements and multiplying by 100. This resulted in a total agreement score of 99.71% (98.43% - 100%).

Procedural Integrity. For 41.9% of all baseline trials, 47.1% of all training trials, 75% of all generalisation trials, and 68.75% of all maintenance trials, a secondary data collector collected procedural integrity data. Two measures of integrity were scored for every trial: trial presentation and correct consequence provided. For each of these categories, either a yes or no was scored. If any item was scored as no, the entire trial was scored as incorrect. The total number of trials scored as correct were divided by the total number of trials scored. This resulted in an integrity score of 98.58% (95.83% - 100%).

Results and Discussion

The initial relational assessment showed that Pt. 4.1 met the criterion for inclusion within the study as she failed the arbitrary hierarchical assessment (see Table 5.3).

Table 5.3. Percentage of correct arbitrary hierarchical relational responses for Pt. 4.1.

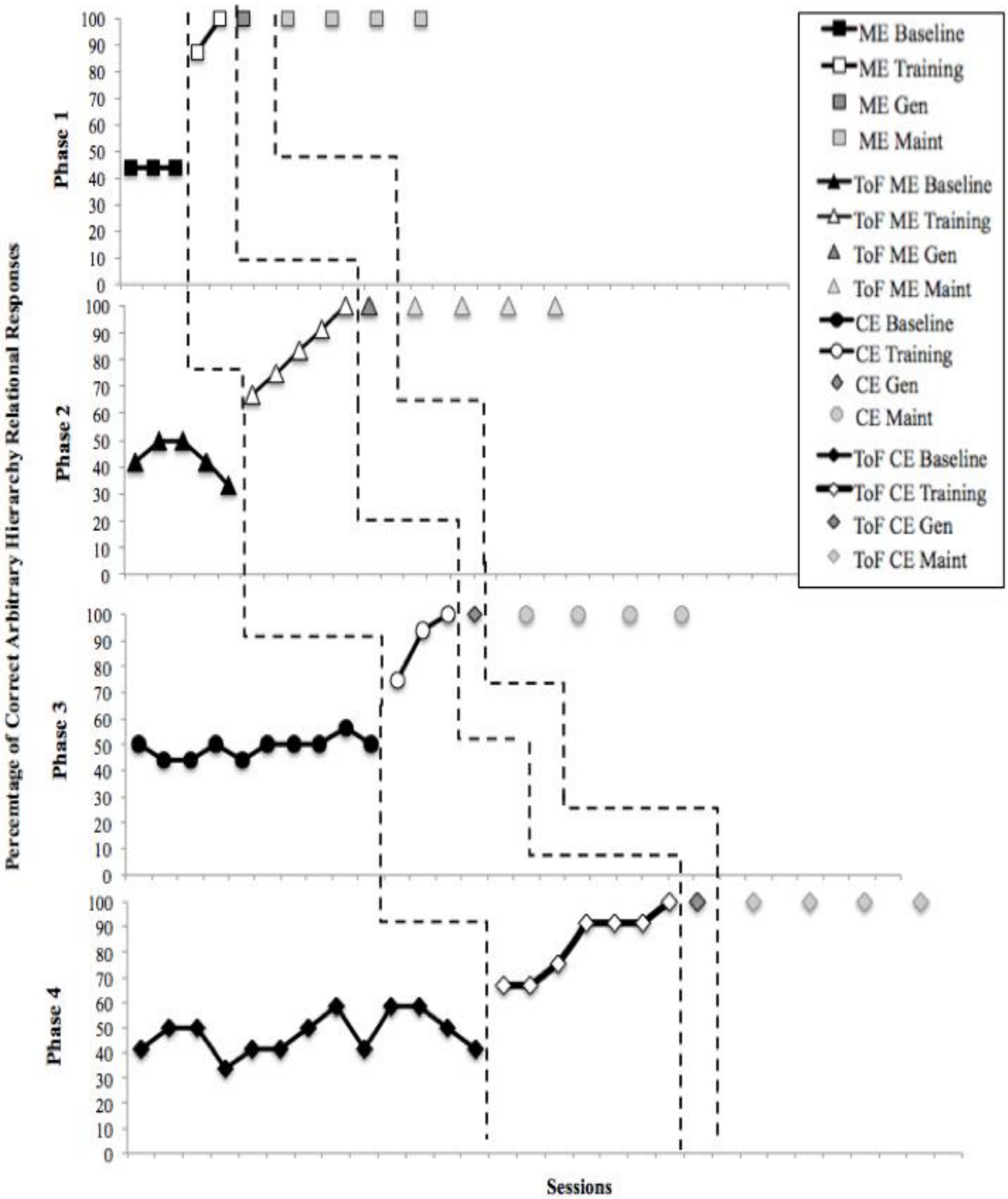
ME	ToF ME	CE	ToF CE	Total Score
43.75%	50%	50%	41.67%	46.42%

Note. ME = Mutual entailment; ToF ME = Transformation of stimulus function of mutually entailed relations; CE = Combinatorial entailment; ToF CE = Transformation of stimulus function of combinatorially entailed relations.

The results of the multiple-baseline analysis are displayed in Figure 5.1. Baseline scores for each relational components remained stable prior to the introduction of the initial phase of arbitrary hierarchical training. The sequential introduction of training for each relational component resulted in increases to criterion for that component. These increases were obtained without corresponding changes in the baseline of the remaining untrained relational components. The results of the multiple baseline analysis for generalisation across untrained stimuli and maintenance effects for trained stimuli are also presented in Figure 5.1. These data demonstrate that she displayed generalisation and maintenance of arbitrary hierarchical relational repertoires.

Pt. 4.1’s baseline arbitrary hierarchical mutual entailment scores remained at 43.75% throughout baseline testing. Following the introduction of arbitrary hierarchy training, her mutual entailment performance increased significantly from baseline levels to 100% over two sessions. She was then exposed to generalisation probes across untrained stimuli after training was terminated.

Figure 5.1. Multiple baseline across components design displaying correct arbitrary hierarchical responses for Pt. 4.1.



Note. ME = Mutual entailment; Gen = Generalisation; Maint = Maintenance; ToF ME = Transformation of function of mutually entailed relations; CE = Combinatorial entailment; ToF CE = Transformation of function of combinatorially entailed relations.

The results of the generalisation probe indicated that she had successfully generalised arbitrary hierarchical mutual entailment responses across untrained stimuli. Maintenance probes were then conducted one week, two weeks, three weeks, and four weeks following the termination of training and also indicated that her training was successfully maintained for up to a month after training was terminated.

Baseline testing indicated that Pt. 4.1's transformation of stimulus function of mutual entailment, consistent with arbitrary hierarchy, varied between 33.33% and 50% correct responding throughout baseline testing, with a downward trend observed for this relational component over the final three baseline sessions. Her performance across this relational component increased significantly from baseline levels to 100% across five sessions. The results of generalisation probes also indicated that the participant had successfully generalised responding across novel stimuli. The results of maintenance probes also indicated that she had successfully maintained training effects for up to one month post-training.

Pt. 4.1's baseline arbitrary hierarchical combinatorial entailment scores remained between 43.75% and 56.25% throughout baseline testing. Her arbitrary hierarchical combinatorial entailment significantly increased from baseline levels to 100% across three sessions once exposed to training. She also demonstrated successful generalisation across untrained stimuli, in addition to maintenance of training effects for up to one month following the cessation of training.

Baseline testing indicated that Pt. 4.1's transformation of stimulus function of combinatorial entailment, consistent with arbitrary hierarchy, varied between 33.33% and 58.33% throughout baseline testing. Following the introduction of arbitrary hierarchy training, her responding increased to 100% across seven training sessions. The results of generalisation probes also indicated that she had successfully generalised responding across novel stimuli. The results of maintenance probes also indicated that she had successfully maintained training effects for up to one month post-training.

The present study indicates that it is possible to establish an arbitrary hierarchical relational repertoire with a typically developing child. Given the dearth of research within this area, the current study offers further insight into the implementation of successful training procedures to facilitate arbitrary hierarchy repertoires and relational responding as a whole, and also adds to the current literature base within this area. The findings of the current study indicate that the combination of arbitrary hierarchy training, contingent feedback and positive reinforcement procedures comprise an effective training method for successfully establishing arbitrary hierarchy repertoires.

Study 5 Introduction

Study 4 demonstrated successful training and testing of arbitrary hierarchical relational responding in a 6-year-old child and indicates that the arbitrary hierarchical training protocol outlined in Study 4 is a potentially useful tool for facilitating this repertoire. The present study (i.e., Study 5) extended the previous research in three ways. Firstly, Study 5 employed single-subject research design that exerted greater experimental control (i.e., combined multiple baseline across people and across arbitrary hierarchical relational components) than that of Study 4. Secondly, Study 5 provided a more comprehensive maintenance assessment than that of Study 4 by assessing arbitrary hierarchical relational responding with the trained stimulus set (i.e., Stimulus Set 1) and with the untrained stimulus sets (i.e., Stimulus Sets 2 – 4), thereby providing information as to whether the participants demonstrated maintenance of both trained and generalised responding up to one month following the termination of the training program. Finally, Study 5 also assessed participants both pre- and post-training for their performance on measures of language and categorisation and compared the training participants' performance on these measures to participants who did not receive training.

Study 5 Method

Participants & Setting

Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland, Galway, Research Ethics Committee. Six typically developing Irish children in the 6- to 7-year-old range all enrolled in a rural Irish primary school served as participants. Consent for conducting the study was obtained from the principal in the school involved. Parental consent was obtained for each child who participated and verbal consent was also obtained from each of the participants.

All six participants were selected based on their performance on an assessment of arbitrary hierarchical relational responding in which they failed (by achieving a score of 60% or less). Three children (i.e., Pt. 5.1, Pt. 5.2, and Pt. 5.3; mean age = 6 years 4.3 months; SD = 2.08 months) were randomly assigned to the training group (i.e., the group that would receive arbitrary containment training. Pt. 5.1 was a 6-year and 6-month-old female, Pt. 5.2 was a 6-year and 5-month-old female, and Pt. 5.3 was a 6-year and 2-month-old female. The remaining children (i.e., Pt. 5.4, Pt. 5.5, and Pt. 5.6; mean age = 6 years 3.3 months; SD =

2.08 months) were designated the non-training group (these children were assessed before and after the training phase but received no training). Pt. 5.4 was a 6-year and 5-month-old male, Pt. 5.5 was a 6-year and 4-month-old male, and Pt. 5.6 was a 6-year and 1-month-old female.

Materials

Participants were exposed to a number of different assessments before the training intervention. These included two standardised measures, namely, the Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn, 2007) and the Children's Category Test (CCT; Boll, 1993); a test of class inclusion (Piaget, 1952) and one test of relational responding, namely, arbitrary hierarchy as previously outlined in Study 4. They received all the same assessments after the training intervention.

PPVT-4. This test assesses receptive vocabulary in respondents aged from 2 years 6 months to late adulthood and is often used as a test of scholastic aptitude. It is administered individually and is presented in a multiple-choice format in which the respondent is presented with four pictures, and asked to select the one that best illustrates the definition of a particular word. The stimuli include items representing up to 20 content areas (e.g., actions, vegetables, tools), and components of speech (nouns, verbs, or attributes), encompassing a broad range of difficulty. Test-retest reliability has been shown to be between .92 and .96 (Community-University Partnership for the Study of Children, Youth, and Families, 2011), while internal consistency has been noted to be similarly high (between .94 and .95). Construct and convergent validity has been assessed by comparing the PPVT-4 to the Expressive Vocabulary Test, Second Edition (EVT-2; Williams, 2007). Correlations between the two were high ($r = .80$ to $.84$) for all age groups (Community-University Partnership for the Study of Children, Youth, and Families, 2011).

CCT. This test was designed to assess categorisation skills for children between 5 and 8 years (Level 1) and between 9 and 16 years (Level 2). It is individually administered and is designed to assess concept formation and problem-solving abilities with novel material. In the current study, only Level 1 was used. This level consists of 5 subtests and includes 80 questions. Within each subtest children are asked to determine the principle underlying correct performance (i.e., matching based on colour, size or proportion) using examiner feedback. Previous research has indicated that the test has adequate test-retest reliability ($r = .75$), in addition to strong internal consistency for both levels 1 ($r = .88$) and 2 ($r = .86$; MacNeill Horton, 1996).

Class Inclusion Test. This is a classic test of categorisation ability that assesses the capacity to respond to a stimulus as simultaneously belonging to both a class and a superordinate class (Thomas & Horton, 1997). It has its origins in Piagetian developmental psychology (Piaget, 1952) where it was seen as determining if a child had reached the “concrete operational” stage of development, hypothesised to occur by age seven or eight. In the test, the child is first shown an array of stimuli, which belong to a particular category. These stimuli form two different subclasses, with a greater number of stimuli within one category than the other (e.g., 5 apples and 2 pears). The child is asked if there are more of the larger subclass or of the category; for example “Are there more apples or are there more fruit?” Materials used in the current study included 5.5cm x 5.5cm colored flashcards depicting a variety of animals (horse, dog, pig, cow, cat, and dog), fruit (strawberry, apple, pear, banana, lemon, orange), clothing (socks, skirt, t-shirt, dress, trousers, coat) and vehicles (car, fire engine, truck, motorbike, tractor, bus). Multiple examples of each of these stimuli were employed.

Arbitrary Hierarchy Assessment. Participants received a test for arbitrary hierarchy, which was identical to that employed in Study 4. Three of the participants were exposed to training in arbitrary hierarchy relations as per Study 4. Participants were assessed both before and after training for all four stimulus sets (i.e., Sets 1 – 4). They were trained using Stimulus Set 1 alone.

Experimental Design

The current study employed a combined multiple baseline design (multiple probe design across behaviours, and a multiple baseline design across participants) with three participants (i.e., Pt. 5.1, Pt. 5.2, Pt. 5.3) (For the benefits associated with each of these experimental designs, see Chapter 4). All participants were exposed to baseline sessions until a stable level of responding was observed. Following this, one participant (i.e., Pt. 5.1) was entered into Phase 1 of training, while the remaining participants remained in baseline conditions. During each phase of training (i.e., mutual entailment, transformation of stimulus function via mutual entailment, combinatorial entailment, and transformation of stimulus function via combinatorial entailment) the remaining untrained relational components were probed during experimental sessions. The two remaining participants were exposed to baseline sessions until the first participant had completed the first phase of training. Thereafter, the next participant (Pt. 5.2) was exposed to training once they had shown stable rates of responding. After that second participant had completed the first phase of training, and again assuming stable responding, then the third participant (Pt. 5.3) was exposed to training.

Procedure

The school provided the setting for the work and all assessment and testing took place during school hours in a classroom separate from the main one in which children's normal lessons took place. Participants were assessed and trained in individual sessions. All participants were initially assessed using the PPVT-4, CCT, and class inclusion test across three sessions ranging in duration between 15 and 30 minutes each. As per Study 4, all participants were assessed for arbitrary hierarchy relations over one session that lasted between 30 – 40 minutes.

Assessing Arbitrary Hierarchy. All participants were assessed for arbitrary hierarchical relations as per Study 4.

Training Arbitrary Hierarchy. The three training participants (Pt. 5.1, Pt. 5.2, Pt. 5.3) were exposed to arbitrary hierarchical training over the course of five weeks. Each training session lasted approximately 40 minutes and sessions were held five times per week. As outlined in the design section, all three participants were exposed to baseline sessions before receiving training in accordance with the sequential logic of the multiple baseline design. The baseline assessment and arbitrary hierarchy training was identical to that described in Study 4.

Generalisation Assessment. Following the completion of each phase of experimental training, the participant was then exposed to generalisation probes using Stimulus Sets 2, 3 and 4 as per Study 4. Generalisation probes were identical in format to baseline assessments and were also comprised of 56 trials. No feedback was provided in these trials. Criterion for demonstration of generalisation was 90% (see Study 3 for rationale regarding generalisation criterion).

Maintenance Assessment. Five weeks following the completion of the fourth phase of experimental training, the participant was exposed to maintenance probes using Stimulus Set 1 (maintenance of training with trained stimulus sets), and Stimulus Sets 2 – 4 (maintenance of training with untrained stimulus sets). As such, any failure to maintain the trained arbitrary hierarchy repertoire could be separated from a failure to maintain the generalised skill. Maintenance probes were identical in format to baseline assessments and were comprised of 56 trials. No feedback was provided in these trials. Criterion for demonstration of maintenance was 90% correct (see Study 3 for rationale regarding generalisation criterion).

Post-Training. All participants (both Training and Non-training) were again assessed for their verbal ability (PPVT-4), classification ability (CCT) and class inclusion responding 6 months after the initial assessments.

Procedural Fidelity and Inter-Observer Agreement (IOA). Sessions were observed in person for IOA by a secondary data collector for 38.49% of all trials. A total agreement score of 99.42% was reached (93.75% - 100%). Procedural fidelity data were collected for 38.49% of all trials. The results indicated that procedural integrity ranged from 97.92% - 100% (average 99.79%).

Results

The initial relational assessment showed that all six participants (i.e., both Training and Non-training) failed the arbitrary hierarchy assessment (see Table 5.4).

The results of the combined multiple baseline design are displayed in Figure 5.2. Baseline scores for the first relational component, mutual entailment, remained stable for all three participants prior to the introduction of the initial phase of arbitrary hierarchical training. The sequential introduction of training (both for mutual entailment as well as for other components) for each participant resulted in increases to criterion for that participant. In the case of each participant, the sequential introduction of training for each relational component resulted in increases to criterion for that component.

These increases were obtained without corresponding changes in the baseline of the remaining untrained relational components and/or participants. The results of the multiple baseline analysis for generalisation across untrained stimuli and maintenance effects for both trained and untrained stimuli are also presented in Figure 5.2. These data show universally successful generalisation and maintenance of the relational repertoires.

As an example of individual data, let us consider the performance of Pt. 5.1. The baseline mutual entailment scores for this participant remained at 10/16 (62.5%) throughout baseline testing. Following the introduction of arbitrary hierarchy training for mutual entailment (Phase 1), this participant's performance increased substantially and met criterion within three training sessions. They then showed successful generalisation to untrained stimuli. Similarly successful outcomes were seen in the case of each of the other relational components (i.e., transformation of function via mutual entailment [requiring 4 training sessions], combinatorial entailment [4 training sessions] and transformation of function via combinatorial entailment [6 training sessions]). This participant also passed maintenance testing, conducted some 5 weeks after the end of the intervention, for each of the relational components and with both trained and untrained stimuli in each case. Similarly successful patterns to those shown by Pt. 5.1 were seen in the case of each of the other two participants (i.e., Pt. 5.2 and Pt. 5.3) also.

Table 5.4. Percentage of Participants' correct baseline arbitrary hierarchical relational responses.

Participant	Mutual Entailment	ToF Mutual Entailment	Combinatorial Entailment	ToF Combinatorial Entailment	Total Score
Training Participants					
Pt. 5.1	62.5%	58.33%	50%	50%	55.21%
Pt. 5.2	56.25%	50%	43.75%	41.67%	47.92%
Pt. 5.3	50%	50%	43.75%	41.67%	46.36%
Comparison Participants					
Pt. 5.4	68.75%	66.67%	50%	50%	58.86%
Pt. 5.5	56.25%	58.33%	50%	50%	53.65%
Pt. 5.6	43.75%	41.67%	43.75%	41.67%	42.71%

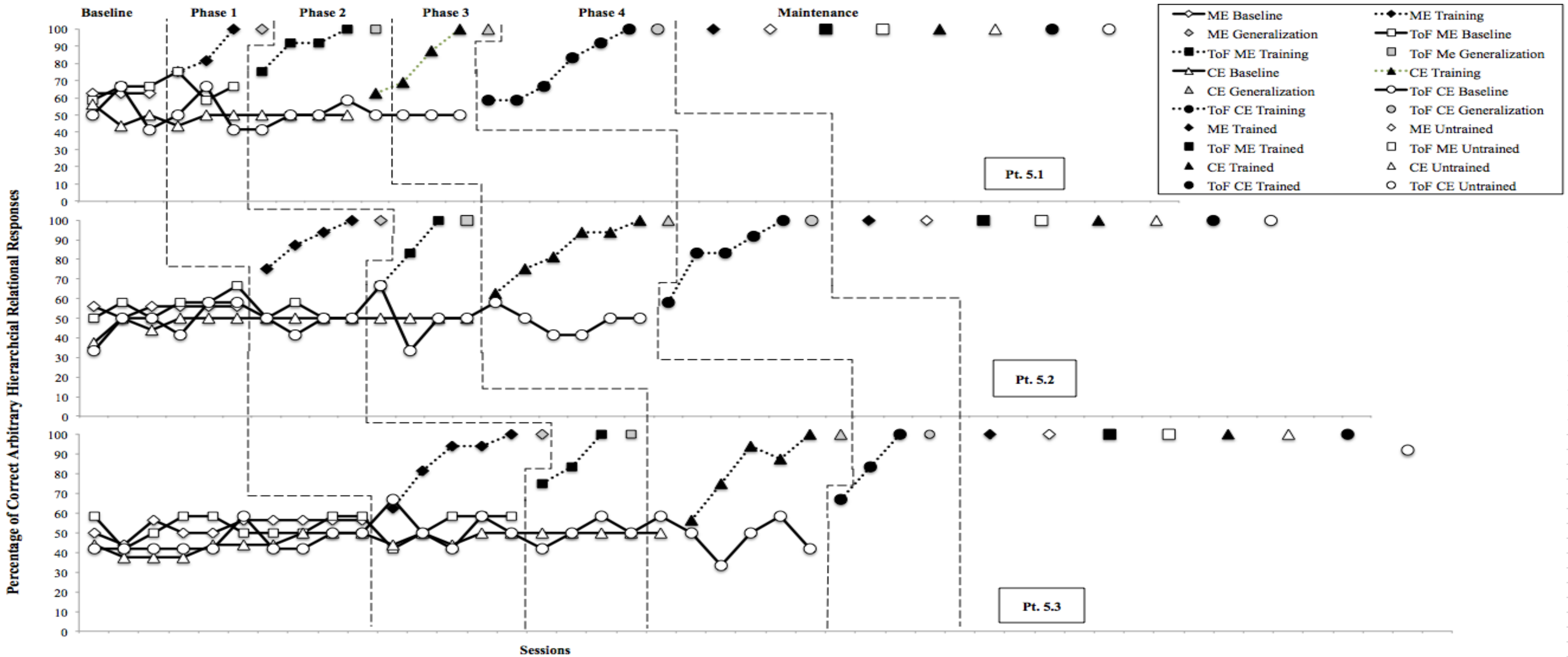
Note. ToF = Transformation of stimulus function.

Both Training and Non-training participants were tested for arbitrary hierarchy repertoires six months following the cessation of training. The results indicated that all Training participants maintained high levels of responding while the Non-training participants showed minimal gains in performance compared to testing at intake (see Table 5.5).

Both Training and Non-training participants were assessed for PPVT-4, CCT and class inclusion scores prior to training, and six months following the cessation of training. The raw scores for each of these measures are displayed in Table 5.6.

The results of this assessment indicated while both the Training and Non-training groups evidenced gains in responding in each of the domains in question, the gains made by the Training group were greater than those of the Non-training group across each of the three instruments. More specifically, the Training and Non-training groups showed mean improvements of 7.3 and 1.7 respectively in the PPVT-4, 2.7 and 0.7 respectively in the CCT and 1.3 and 0.3 respectively in the class inclusion test.

Figure 5.2. Combined multiple baseline across components and across participants design displaying correct arbitrary hierarchical responses for Pts. 5.1, 5.2, and 5.3.



Note. ME = Mutual entailment; ToF ME = Transformation of function of mutually entailed relations; CE = Combinatorial entailment; ToF CE = Transformation of function of combinatorially entailed relations.

Table 5.5. Correct arbitrary hierarchy scores at intake and six months post-training.

Participants	ME		ToF ME		CE		ToF CE		Total Score	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
Training Participants										
Pt. 5.1	62.5%	100%	58.33%	100%	50%	100%	50%	100%	53.13%	100%
Pt. 5.2	56.25%	100%	50%	100%	43.75%	100%	41.67%	100%	46.88%	100%
Pt. 5.3	50%	100%	50%	100%	43.75%	100%	41.67%	100%	43.75%	100%
Comparison Participants										
Pt. 5.4	68.75%	68.75%	66.67%	66.67%	50%	50%	50%	50%	54.69%	58.86%
Pt. 5.5	56.25%	62.5%	58.33%	58.33%	50%	50%	50%	50%	48.44%	55.21%
Pt. 5.6	43.75%	50%	41.67%	50%	43.75%	43.75%	41.67%	50%	42.19%	48.44%

Note. ME = Mutual entailment; ToF = Transformation of function; CE = Combinatorial entailment; T1 = Time 1 (intake score); T2 = Time 2 (six months-post training).

Table 5.6. Pre-Training Raw scores of Training and Comparison Participants.

Participant	PPVT-4 Raw Score			CCT Raw Score			CI Score		
	T1	T2	Score Change	T1	T2	Score Change	T1	T2	Score Change
Training Participants									
Pt. 5.1	108	116	+8	66	69	+3	6	7	+1
Pt. 5.2	107	114	+7	69	71	+2	5	7	+2
Pt. 5.3	100	107	+7	62	65	+3	5	6	+1
Comparison Participants									
Pt. 5.4	104	106	+2	67	68	+1	5	6	+1
Pt. 5.5	104	105	+1	62	63	+1	5	5	0
Pt. 5.6	98	100	+2	64	64	0	5	5	0

Note. PPVT-4 Raw Score = Peabody Picture Vocabulary Test, 4th Edition raw score; CCT Raw Score = Children's Category Test raw score; T1 = Time 1 (intake score); T2 = Time 2 (six months-post training).

Discussion

Studies 4 and 5 provide evidence that hierarchical relational framing repertoires, argued by RFT as being central to classification responding, may be successfully taught to young children. In Study 4, one six-year-old girl was successfully taught arbitrary hierarchical relational framing in the context of a multiple baseline across components design, while Study 5 was concerned with training three further children in the 6- to 7-year-old age range in arbitrary hierarchical relational framing using a multiple baseline design across both components of the relational framing repertoire and across participants.

The data generated from Studies 4 and 5 contribute to previous empirical RFT research showing that repertoires of relational framing can be taught using MET (e.g., Barnes-Holmes, Barnes-Holmes & Smeets, 2004; Barnes-Holmes, Barnes-Holmes, Smeets, Strand & Friman, 2004; Berens & Hayes, 2007) and also extend the research reported in Chapter 4 (i.e., Studies 2 and 3). As in Berens and Hayes (2007) and Study 3 within the current thesis, Study 5 further contributes to the extant literature by examining and training arbitrarily applicable hierarchical relational responding in the context of a combined multiple baseline design (i.e., across relational components and across participants). Both Studies 4 and 5 have provided further evidence of the operant nature of relational framing and its components (i.e., mutual entailment, combinatorial entailment and transformation of stimulus functions). These studies also demonstrate acquisition and generalisation of the hierarchical relational framing repertoires, and also show a high level of maintenance of these repertoires over time. These studies indicated a five-week maintenance of arbitrary hierarchical responding for both trained (Studies 4 and 5) and untrained (Study 5) stimulus sets. This positive maintenance outcome provides further indication that MET constitutes an effective procedure for establishing relational framing repertoires.

Apart from showing the successful training of relational framing relevant to classification itself (i.e., hierarchy), Study 5 also provides evidence that suggests that training such framing can boost intellectual performance. More specifically, in Study 5 the children who were trained in arbitrary hierarchical relational framing showed improvements in their performance on standardised tests of linguistic and cognitive ability, including an assessment which targeted classification responding (i.e., the CCT). The results of Study 5 indicate that this pattern of improved performance across these assessments was markedly better than that seen in the case of a control group of children who had not received training. The general pattern in evidence here is similar to that seen in other RFT studies which have shown that relational frame training can boost general intellectual performance (see for example, Cassidy, Roche & Hayes, 2011; Stewart & Hayes, 2016). Although the current data are based on particularly small numbers of participants; nevertheless, they are promising and indicative for further work involving larger numbers.

In addition to this, the results of Study 5 indicated that it was not simply general intellectual performance which seemed to be improved by the hierarchical framing training provided but performance on tests of classification. The latter were administered because the framing repertoires taught were assumed to be relevant to classification as suggested by the results of Study 2 contained in Chapter 3. Of course, it is possible that simply training relational framing of any kind might boost performance on any test of cognitive potential, including ones specific to classification. Accordingly, future research will need to explore whether particular repertoires of framing might be more likely than others to improve performance within particular intellectual domains and more specifically, whether training hierarchical framing might support classification in particular, above and beyond other skills.

These findings have considerable implications in terms of applied work with various populations, including both typically developing children as well as populations with developmental delay. One of the advantages of RFT is that it identifies an exact unit to target that is regarded as fundamental to human language and cognition. Knowing the unit, or relational frame, that is central to particular behavioural repertoires (e.g., deictic relational framing is fundamental to the execution of perspective-taking) is an important consideration when training these repertoires. Hayes and colleagues (2001, p. 28) argued that it was critical to determine “whether we can design effective RFT-based interventions that establish or facilitate new repertoires of derived relational responding in young children”. The present research supports the efficacy of an RFT-based intervention to establish derived relational responding repertoires consistent with hierarchy among young children. Given the growing body of data on the link between relational behaviour and language and cognitive abilities (e.g., Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; Mulhern, Stewart & McElwee, 2017), this result supplements previous work and provides yet further potential tools to the practitioner interested in establishing core intellectual skills in children or other populations.

The success of the current protocol provides a platform for further work examining relational framing of hierarchy and classification. Such work might address potential limitations of the paradigm used here as well as advancing the agenda of the investigation of this domain. One obvious limitation already mentioned is in respect of the capacity to investigate the effects of the training provided in the context of Study 5 on intellectual performance. Due to the very small sample sizes used, the current data cannot yield definitive conclusions in this respect either regarding improvement in general intellectual performance or specifically with regard to possible improvement of the repertoire of categorisation or classification. The small sample size used in both Studies 4 and 5 was appropriate for the multiple baseline design, which was used to show that MET could establish pattern of hierarchical relational framing with the 6- to 7-year-old children. However, in regards to Study 5 the sample size employed was too small to allow conclusions to be drawn with regard to the comparison between the experimental and control groups. Future work could contribute by deliberately targeting the latter instead, by employing a similar training protocol as used in Studies 4 and 5 respectively, but in the context of a large between-groups design.

Such work might for example, assess three different groups before and after hierarchical training: 1) a control group to whom no training is provided, 2) a hierarchical training group, and 3) a relational frame training group establishing an alternative relational frame for whom the quantity of relational frame training received is yoked to the quantity received by the second group, but whose training is in some other pattern of framing. Such research could provide greater insight into the effect on intellectual performance of hierarchical relational frame training by allowing comparisons between Group 2 and each of the other two groups. One might predict that both the relational frame training groups might demonstrate improvements in intellectual performance compared with the control group. However, it is also possible that the two relational framing groups might differ in this respect or, possibly more interestingly, that the hierarchical framing group might outperform the other group specifically in terms of categorisation and classification performance.

Another potential limitation of the current work is that it might be argued that the hierarchical training protocol did not provide sufficiently thorough or comprehensive training of the repertoire of hierarchical framing as might be desirable. For example, while framing appropriately in response to the cues of “type of” and “contains” is an important part of this pattern, there are other aspects of this pattern that might also be tested. For instance, when two stimuli are framed as being part of the same class, then they should also both be framed as being different from other stimuli framed within a different class and equivalent to each other, in at least some contexts, independent of their physical properties. In addition, and related to this, functions trained to one particular stimulus in a hierarchical relational network might be expected not only to show transfer in particular directions up and down the network but also (selectively) across the network – for example, strongly to stimuli in the same class within the hierarchy while much more weakly if at all to stimuli in other classes at the same hierarchical level. Such a pattern of responding might be expected from someone with a sufficiently advanced repertoire of hierarchical classification. The current protocol did not involve testing or training such relations. Adjusting it so as to do so might be a useful direction for future work.

In summary, the current findings suggest that hierarchical relational framing can be successfully trained in young children in whom such repertoires are absent or deficient. In addition, there is some indication that training such a repertoire can boost intellectual ability. Further work, however, is needed to provide a more thorough and comprehensive training of the relational framing that underlies classification as well as to provide more convincing evidence as to the effects of training this repertoire on intellectual performance in general as well as specifically with regard to classification.

Chapter 6

Studies 6 - 8

Teaching Class Inclusion to Typically Developing Children and Individuals with Autism*

* Portions of this chapter have been accepted for publication:

Ming, S., Mulhern, T., Stewart, I., Moran, L., & Bynum, K. (in press). Training class inclusion responding in typically-developing children and individuals with autism. *Journal of Applied Behavior Analysis*

* Portions of the material in this chapter are in preparation for journal submission as:

Zagrabska, P., Mulhern, T., & Stewart, I. (in preparation). Testing and training class inclusion responding based on a relational frame theory approach with individuals with autism.

The research presented in the current chapter focused on assessing and training children in 'class inclusion' responding, which involves responding to stimuli as being simultaneously part of both a class and a superordinate class and which can be taken as showing responding in accordance with hierarchical classification.

A class inclusion task has been used in several of the studies in the current research programme as one of the additional measures of intellectual performance to be correlated with performance in accordance with the relational repertoires seen within Relational Frame Theory (RFT) as associated with hierarchical classification (i.e., non-arbitrary and arbitrary containment and arbitrary hierarchy). The findings of Study 1 (Chapter 3) indicated that class inclusion performance was moderately correlated with these repertoires. Studies 3 and 5 (Chapters 4 and 5 respectively) evaluated the impact of training the key relational frames on other intellectual repertoires including class inclusion responding, and suggested that the training could enhance class inclusion performance. In the final phase of the current doctoral programme, I decided to investigate class inclusion responding per se in both typically developing children and children with Autism Spectrum Disorder (ASD) and to investigate the utility of an RFT-based procedure to facilitate this pattern of responding. Before describing the studies, this chapter will first provide a brief review of relevant mainstream work on class inclusion.

Class Inclusion - Background

A typical class inclusion type task (see e.g., Thomas & Horton, 1997) might involve being presented with, for example, five dogs and three cats and being asked, "Are there more dogs or more animals?" Answering correctly that there are more animals than dogs involves responding to a stimulus (e.g., dog) simultaneously as both a member of a class (i.e., of dog) and superordinate class (i.e., of animals). Acquiring this skill has been recognised as important not just by psychologists but also within the educational domain. For example, it has been included in the primary school math curriculum and is also necessary for algebraic grouping (California Department of Education, 2008; Canobi, Reeve & Pattison, 2002; Cowan & Renton, 1996; Langford, 1981).

Developmental psychologists have used class inclusion as a marker for a child's developmental stage (i.e., the Piagetian stage of concrete operations, thought to start around the age of 7) as well as more specifically as a reflection of their understanding of hierarchical categorisation. Winer's (1980) review suggested that class inclusion typically emerges later than age seven, as originally proposed by Piaget. Winer states that the 75% criterion is generally not reached until age ten (Ahr & Youniss, 1970; Carson & Abrahamson, 1976; Kofsky, 1966; Lovell, Mitchell, & Everett, 1962; Meadows, 1977) and often it emerges even later (Billow, 1975; Hooper, Sipple, Goldman & Swinton, 1979). Winer (1980) also maintained that there were grounds to speculate that the addition of specific verbal cues, presented with relational terminology, influence the class inclusion performance of children.

Class inclusion training with typically developing populations. Several researchers have assessed the practicality of training young children in class inclusion (Agnoli, 1991; Brainerd, 1974; Gash, 1982; Goswami & Pauen, 2005; Greene, 1991; Hendricks, Pasnak, Willson-Quayle, Trueblood, Malabonga & Ciancio, 1999; Hooper, Wanska, Peterson & de Frain, 1979; Judd & Mervis, 1979; McCabe & Siegel, 1987; Pasnak, Cooke & Hendricks, 2006; Pellegrini, 1983; Schwebel & Schwebel, 1974; Siegler & Svetina, 2008; Siegel, McCabe, Brand & Matthews, 1978; Thomas & Horton, 1997; Wilson-Quayle & Pasnak, 1997; Wohlhill, 1968). Judd and Mervis (1979) assessed the effect of class inclusion training, including quantification and feedback, with five-year-old children. Participants were divided into three groups. In the first group, participants counted the sum of objects in both the superordinate and subordinate groups. The researcher then delivered verbal feedback regarding the accuracy of their responses. A memory aid was also given to the participants - a card displaying a number matching the sum of objects just tallied. The first group was also asked to provide justifications for their responses. The second group was asked only to count, but received no feedback, while the third group served as controls. The first group showed the most significant gains in class inclusion responses, while responses from the second and third groups remained poor. As such, Judd and Mervis indicate that class inclusion with a population of five-year-old children is feasible if the appropriate training procedures are put in place.

Brainerd (1974) determined the efficacy of an intervention involving the provision of verbal feedback contingent on participants' correct responses of class inclusion tasks with children aged four and five. This treatment was compared with a control group who did not receive verbal feedback. The intervention was found to be successful at increasing the children's class inclusion accuracy. Furthermore, the treatment effects were found to generalise to alternative and untested class inclusion stimulus sets. Several more studies have also successfully employed feedback of various types to augment class inclusion understanding with children (Ahr & Youniss, 1970; Judd & Mervis, 1979; Siegler & Svetina, 2006). However, while Brainerd stated that maintenance of class inclusion was tested, Brainerd only tested class inclusion maintenance one week after training cessation. As such, maintenance cannot be inferred from this.

The premise of Pellegrini's 1983 research is that the ability to observe parallels between subordinates approximates class addition, while the ability to observe discrepancies between subordinates and the superordinate is similar to the skill of class subtraction. As such, the training used by Pellegrini attempted to increase participants' discrimination of similarities and differences between subordinates. Class inclusion training was completed with 160 participants, between seven- and ten-years-old. A series of open-ended questions, hypothesised to aid class inclusion formation, were presented in a specified order. This training substantially increased the participants' class inclusion accuracy. Furthermore, the results indicated that the children in the higher age groups obtained class inclusion at a superior rate.

McCabe and Siegel's 1987 research with kindergarten-aged participants showed that while class inclusion training proved to be initially successful, gains dissipated within three months. The training involved the provision of verbal feedback to participants; this was compared against two control groups. Initially, post-tests showed superior outcomes for the group receiving training. However, a second post-test three months later revealed no significant difference between groups.

Thus, overall, while some success has been achieved in training class inclusion in children even as young as four, there remains a lack of evidence on generalisation and maintenance of training effects, with few studies assessing either generalisation (Brainerd, 1974; Pellegrini, 1983; Wohlhill, 1968) or maintenance (Agnoli, 1991; Hooper et al., 1979; McCabe & Siegel, 1987;

Wilson-Quayle & Pasnak, 1997) effects. Of those which have, one reported positive generalisation outcomes (Brainerd, 1974), while two others reported mixed outcomes (Pellegrini, 1983; Wohlwill 1968). Of the studies that examined maintenance, one reported positive outcomes (Wilson-Quayle & Pasnak, 1997), while another reported mixed outcomes (McCabe & Siegel, 1987). As such, further work is needed to investigate potential training procedures especially regarding maintenance and generalisation.

Class Inclusion and Relational Frame Theory

From the RFT perspective, class inclusion involves both non-arbitrary relational responding (i.e., comparison between groups of physically different size/quantity) and arbitrary relational responding (i.e., containment and hierarchical relations between members and classes and between classes themselves). From this point of view, class inclusion is a subset of advanced hierarchical categorisation and an important step towards the acquisition of this repertoire.

The series of studies presented in this chapter aimed to investigate whether an RFT approach might be useful in teaching class inclusion reasoning to very young typically developing children as well as developmentally delayed individuals. This work will supplement the previous studies in the current thesis as well as contributing to behaviour analytic research with young children facilitating early emergent derived relations supporting hierarchical classification. One other example of a behaviour analytic and more specifically RFT study relevant in this regard was focused on strengthening framing repertoires under the contextual cues of “same” and “different” of both typically developing children and children with developmental disabilities (Newsome, Berens, Ghezzi, Aninao,

& Newsome, 2014). While this study was informative and the data no doubt useful to practitioners, there is a dearth of such work more generally.

The studies presented in this chapter aimed to help address this gap in the literature by assessing and establishing class inclusion responding in young children, both typically developing and diagnosed with ASD, using RFT procedures. The first of these studies, Study 6, aimed to train class inclusion responding in typically developing 3- to 4-year-olds using a protocol incorporating non-arbitrary relational support. As part of this, it examined the generalisability of the target repertoire as well as short-term maintenance (i.e., one month). Study 7 aimed to establish class inclusion repertoires in individuals with a diagnosis of autism using the same basic RFT-inspired protocol. Study 8 extended Study 7 via several modifications, including changes to experimental design, training procedure and generalisation assessments.

Study 6

Method

Participants. Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland, Galway, Research Ethics Committee. Consent for conducting the study was obtained from the manager of the playschool. Parental consent was obtained for each child who participated and verbal assent was also obtained from each of the participants. Three typically developing children between the ages of 3.4 and 4.1 years (*M* = 3.7 years) served as participants. The participant included two males (Pt. 6.1 = 3 years 5 months; Pt. 6.3 = 4 years 1 month) and one female (Pt. 6.2 = 3 years 6 months). Following data collection and analysis, debriefing reports were issued to both the school and parents of the participants. Participants received verbal feedback regarding their performance throughout the research and a certificate of participation.

Setting. The research was carried out in a rural playschool that catered for 25 children, all typically developing. Baseline and non-arbitrary class inclusion training were conducted in a separate classroom adjacent to the participants' main classroom.

Experimental Design. A multiple baseline design across participants was employed in which participants were pre-assigned to one of three baseline lengths (3, 5, or 7 sessions), as advised by Watson and Workman (1981) (See Chapter 4 for further information regarding multiple baseline designs across participants). Previous pilot research, also conducted with typically developing children (Mulhern & Stewart, 2014), indicated that three baseline sessions was adequate to show a stable trend. This research also indicated that an increase in class inclusion responding was likely to be demonstrated soon after the introduction of training. Both of these factors are important in relation to the execution of a multiple baseline design (Christ, 2007).

Materials. The materials included colored flashcards (5.5cm x 5.5cm) of various animals (horse, dog, pig, cow, cat, and dog), fruit (strawberry, apple, pear, banana, lemon, orange), clothes (socks, skirt, t-shirt, dress, trousers, coat) and vehicles (car, fire engine, truck, motorbike, tractor, bus). Multiple examples of each of these subordinate stimuli (i.e., multiple examples of cats, strawberries, socks, cars, etc.) were used. In addition, a number of plastic boxes were employed, one large box with the superordinate category name (i.e., animals) on the front and two smaller boxes with one velcroed slot to which 8cm x 8 cm flashcards with category names (e.g. horse, cow, etc.) could be attached.

Tokens served as generalised conditioned reinforcers, which participants could exchange on a Fixed Ratio (FR) 4 schedule for a choice of edibles, stickers, markers, pens, pencils and erasers. Participants also received positive reinforcement in the form of praise contingent upon correct responses.

Procedure

Screening Assessment

Each participant was first screened for their suitability for the current study prior to their participation. Participants were tested via table top discrete trial procedures to assess their ability to (1) tact all stimuli within the current study, (2) correctly respond to yes and no stimulus identification questions (e.g., “Is this an apple?”), (3) tact the category of the stimuli within the study, (4) tact quantities of the stimuli from one to ten, and (5) correctly respond to non-arbitrary relation questions (e.g., “Are there more horses or more sheep?”).

Baseline

The baseline consisted of providing participants with a choice of stimulus selection cards, which would dictate the quantity of each subclass to present (e.g., 2 tractors and 5 motorbikes). Participants were then presented with a series of questions, with some serving as distractors (e.g., “Are there more dogs than cats?”; “How many cats are there?”), while others targeted class inclusion responding. These included 8 questions that were subsequently divided into “more” or “less” questions (See Table 6.1 for class inclusion questions).

Table 6.1. Class inclusion questions: Eight variations of more and less, counterbalanced for order of stimulus type and category.

More [category] or more [smaller subclass]?	More [category] or more [bigger subclass]?
Less [category] or less [smaller subclass]?	Less [category] or less [bigger subclass]?
More [smaller subclass] or more [category]?	More [bigger subclass] or more [category]?
Less [smaller subclass] or more [category]?	Less [bigger subclass] or less [category]?

The presentation of these questions was randomly selected via the shuffling of the question cards. The remaining 8 questions also focused upon “more” or “less” relations between the subclass stimuli (e.g., “are there more pears or more strawberries?”), in addition to questions of quantity (e.g., “How many dogs are there?”), and questions of features of stimulus class (e.g., “Which of these barks?”). Only questions focused upon class inclusion reasoning were pertinent to the current study; however, both class inclusion responses and distractor responses were recorded. In total eight trials were conducted for baseline assessment of class inclusion, with an additional eight distractor trials. A session was terminated after all eight types of class inclusion questions had been asked. Participants received non-contingent reinforcement for their participation. Baseline sessions typically lasted 10 minutes, and were conducted 5 times per week.

Intervention

Class inclusion multiple exemplar training (MET) was delivered across two phases. The first phase involved pre-trial requirements designed to emphasise the salience of the boxes and the containment relation. The second phase was designed to fade these pre-trial requirements (See Appendix D for script). One large box that contained two smaller plastic boxes (see Figure 6.1) was employed to increase the salience of the containment relation. Only the “animals” stimulus set was employed within the training phase, but the trial types and randomisation of trials remained the same as within the baseline phase.

Figure 6.1. Nested plastic boxes.

Phase 1. Following baseline assessment, the first phase of non-arbitrary training was conducted using the animals stimulus set. The non-arbitrary training involved intervening with participants on a 1:1 basis to teach class inclusion using the Piagetian class inclusion task outlined in the baseline sessions. Stimulus selection and class inclusion questions were again selected by shuffling cards. A large transparent plastic box, for the category, was placed in front of the participants in addition to two smaller transparent plastic boxes, which had a Velcro slot attached for the stimulus types. When presented with the boxes, participants were then told, “This big box is for the category. What category do these belong to?” Once the participants correctly responded to this question, they

were then told, “That’s right. Let’s attach the category label of ‘animals’ to the box”. Following this, the participant’s attention is then drawn to the two smaller boxes and are informed “These two smaller boxes are for the different types of animals. The small boxes will go inside the animal category box”. Participants were then told to attach labels to the small boxes (i.e., “let’s attach the label of ‘dog’ to this small box. Let’s attach the label of ‘cat’ to the other small box”). Participants were then asked to put the stimuli into their respective boxes (i.e., “Put the dog into the dog box, and put the cat into the cat box”).

Once the stimuli were placed within the appropriate boxes, participants were then told, “Great, dogs and cats are types of animals, they belong to the animal category, so they all go inside the big animal category box”. Participants were then asked to identify the boxes based upon the question trial type selected (i.e., if the card for “more [bigger subclass] or more [category]?” was selected, participants were then asked to identify the [bigger subclass] box, and the [category] box). Incorrect responses within this phase were corrected via modelling procedures and a re-presentation of the question. Participants were then presented with the class inclusion question, while the experimenter simultaneously picked up the appropriate boxes (i.e., “Are there more [bigger subclass] (picks up bigger subclass box), or more [category] (picks up category box)?”). Correct responses were consequated with token and praise while the experimenter simultaneously picked up the appropriate boxes - “That’s right, there are more animals (picks up box) than dogs (picks up box)” for “more” questions and “That’s right there are less dogs (picks up box) than animals (pick box)” for “less” questions.

Following incorrect responses, participants were then required to again correctly identify the appropriate boxes. The experimenter then picked up the boxes and stated “That’s right, [stimulus 1 / 2] and [stimulus 1 / 2] are types of animals, so they all go inside the big animal category box. They all belong to the animal category, but only these are [stimulus 1 / 2], so there are more animals in the animal category box than there are [stimulus 1 / 2] in the [stimulus 1 / 2] box” for “more” questions, or “That’s right, [stimulus 1 / 2] and [stimulus 1 / 2] are types of animals, so they all go inside the big animal category box. They all belong to the animal category, but only these are [stimulus 1 / 2], so there are less [stimulus 1 / 2] in the [stimulus 1 / 2] box than there are animals in the animal category box” for “less” questions. Participants were then re-presented with the class inclusion question. Once participants provided a correct response, a new stimulus set was chosen and participants were presented with the same class inclusion question again. Specific praise and tokens were provided to participants for their correct responses. Only trials with first stimulus sets were recorded and included within the graphic analysis. Non-class inclusion questions were also recorded. The criterion to move onto the next phase of training was to get 8 out of 8 correct in one session.

Phase 2

The second phase of training was identical to the previous phase of training. However, (a) the requirement to identify the boxes and (b) the verbal reference to the size of the boxes were eliminated from the trials. Stimulus selection and class inclusion questions were again selected by shuffling cards. A large transparent plastic box, for the category, was placed in front of the participants in addition to two smaller transparent plastic boxes, which had a

Velcro slot attached for the stimulus types. Participants were then asked to put the stimuli into their respective boxes (i.e., “Put the pigs into the pig box, and put the cow into the cow box”), and place the smaller boxes into the larger box (i.e., “Now put all of them inside the animal category box”). Participants were then presented with the class inclusion question; however, in this phase the experimenter did not pick up the boxes when presenting the question. Correct responses were reinforced with both tokens and praise while simultaneously picking up the appropriate boxes - “That’s right, there are more animals (picks up box) than dogs (picks up box)” for “more” questions and “That’s right there are less dogs (picks up box) than animals (pick box)” for “less” questions.

For incorrect responses, participants were then required to again correctly identify the appropriate boxes. The experimenter then picked up the boxes and stated “They all belong to the animal category, but only these are [stimulus 1 / 2], so there are more animals than there are [stimulus 1 / 2]” for “more” questions, or “They all belong to the animal category, but only these are [stimulus 1 /2], so there are less [stimulus 1 / 2] than there are animals” for “less” questions. Participants were then re-presented with the class inclusion question. Once the participant had provided a correct response, a new stimulus set was chosen, and they were again presented with the same class inclusion question again. Specific praise and tokens were provided to participants for their correct responses. Only trials with first stimulus sets were recorded and included in the graphic analysis. Non-class inclusion questions were also recorded. The criterion to complete training was 8 out of 8 correct in one session. This mastery criterion was chosen as a rate of 100% correct responding was necessary to ensure mastery given the the small number of trials presented to the participant (i.e., 8).

Generalisation and Maintenance

Generalisation testing was identical to that of baseline. However, the first phase of generalisation testing involved only the presentation of the stimulus set of animals, while the second phase of generalisation testing involved testing novel categories, such as vehicles, fruit and clothing. Maintenance testing was conducted 1 month following each participant's acquisition of correct class inclusion responding. Maintenance testing was identical to that of baseline. If the maintenance probe yielded results above 75% class inclusion responding was deemed as maintained. This rate was selected based upon research by Winer (1980) who suggested that a sufficiently strong class inclusion repertoire is present when an individual demonstrates correct responses for 75% or more of class inclusion tasks.

Inter-Observer Agreement (IOA) and Procedural Fidelity

IOA was conducted for all sessions, including baseline, intervention, generalisation and maintenance sessions. IOA was conducted by a research assistant and was also conducted within the participants' school. Prior to data collection, IOA collectors were trained in class inclusion data collection until they reached 100% accuracy. Mean IOA for this study was 99% (Range 87.5 – 100%). Procedural fidelity was conducted for all sessions, including baseline, intervention, generalisation and maintenance sessions. A research assistant collected procedural fidelity measures. Prior to data collection, correct experimenter procedural responding was modelled to the data collectors for exemplars of procedural integrity, and data collectors were also provided with a printed copy of the training protocol and a checklist of the steps within the procedure. Mean procedural integrity for this research was 98.6% (Range 87.5 - 100%).

Results and Discussion

The results of the multiple-baseline design are shown in Figure 6.2. Baseline scores for all participants remained stable prior to the introduction of non-arbitrary class inclusion training. In the case of each participant, the introduction of class inclusion training resulted in a considerable increase over baseline levels. These changes were obtained without corresponding changes in the baselines of the remaining untreated participants. The results for generalisation and maintenance are also presented in Figure 6.2. These data demonstrate that each participant displayed generalisation and maintenance of the trained repertoire (i.e., class inclusion responding).

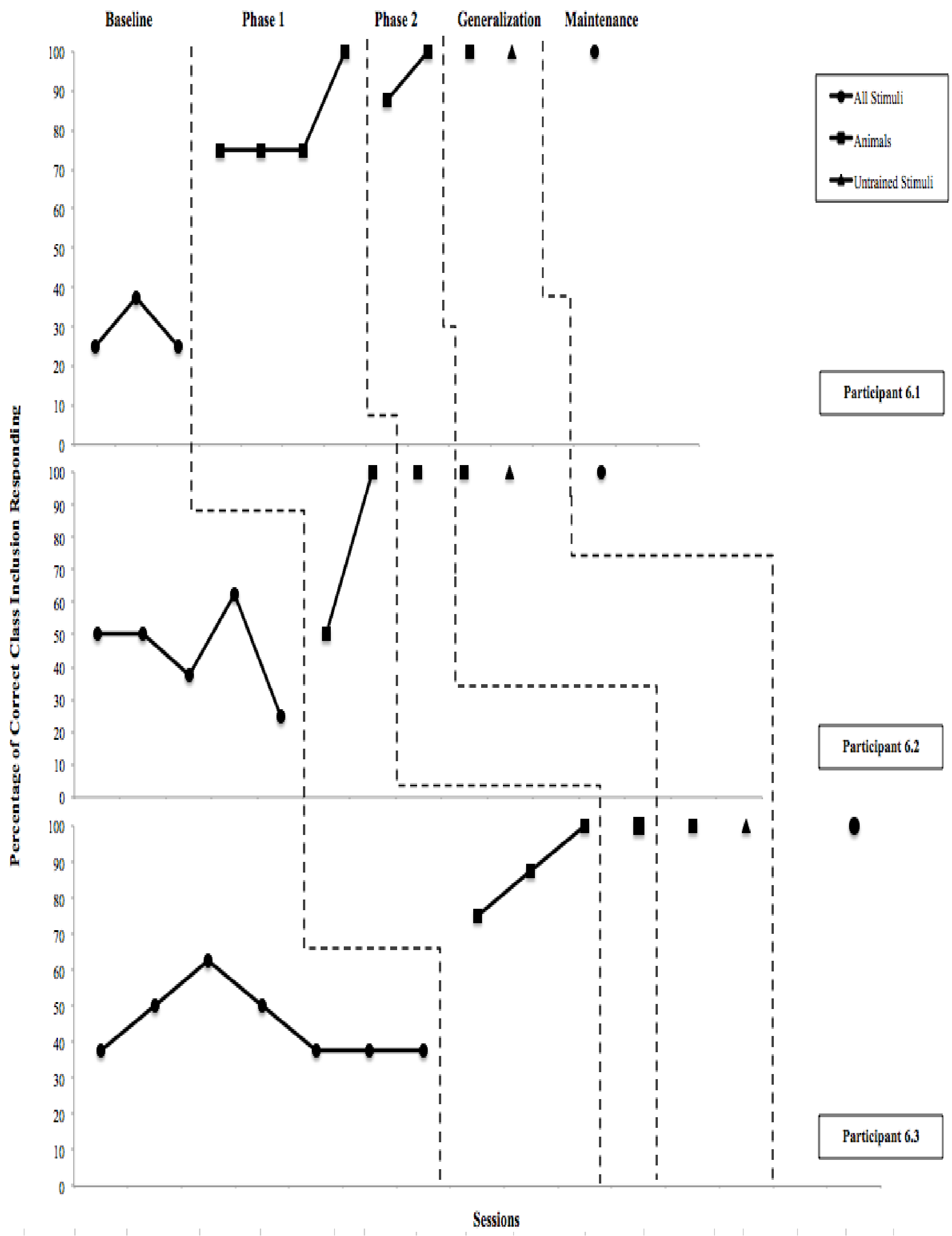
Pt. 6.1's baseline class inclusion remained between 25 and 37.5% throughout baseline testing. Pt. 6.1's class inclusion performance increased from this baseline level to 100% over four sessions within the first phase of class inclusion training. The class inclusion responses for Pt. 6.1 remained high for the second phase of training, with the participant passing following one training session. Following class inclusion training termination, Pt. 6.1's generalisation of class inclusion with the stimulus set of animals was tested and found to have met generalisation criteria (i.e., 100% correct). An additional generalisation test using novel categories also showed that the participant met generalisation criteria with these untrained categories. Maintenance probes were then conducted one month following training termination, and the results indicated that the participant met maintenance criteria (i.e., 100% correct).

Pt. 6.2's baseline class inclusion performance ranged between 25 and 62.5%, but was stable throughout baseline testing. Pt. 6.2's class inclusion performance increased to 100% following two sessions of the first phase of class inclusion training. The second phase of training was similarly high, with the participant attaining 100% correct responding following one session. Pt. 6.2 also demonstrated generalisation across both generalisation probes and also met mastery criterion at one-month follow-up.

Pt. 6.3's baseline class inclusion performance ranged from 37.5 to 62.5%, but was stable throughout baseline testing. Following the initial phase of training, Pt. 6.3 met criteria following three training sessions. Pt. 6.3 maintained high levels of responding within the second phase of training, meeting training criterion after one session. Pt. 6.3 also showed generalisation and maintained class inclusion responding at one-month follow-up.

Findings were that this study successfully trained class inclusion responding in children between three- and four-years using a controlled (multiple baseline) design as well as showing maintenance and generalisation. Training of this repertoire in this age group is unprecedented in published literature. This is a success for a behaviour analytic and more specifically RFT-based approach to training class inclusion.

Figure 6.2. Correct Class Inclusion Responses of Three- to Four-Year-Olds.



Baseline results indicated low levels of class inclusion responding, between 25 and 62.5%, as would be expected within this age group (3-4 years) based on past theory and research (e.g., Inhelder and Piaget, 1964). At the same time, in accordance with behaviour analysis and RFT, MET of the pattern was shown to facilitate the early emergence as well as both generalisation and maintenance of this repertoire.

Previous theorists have debated the initial age of class inclusion development (Brainerd, 1974; Hooper, Sipple, Goldman & Swinton, 1979; Meadows, 1977; Winer, 1980) as well as the efficacy of training (Goswami & Pauen, 2005; McCabe & Siegel, 1987; Pasnak, Cooke & Hendricks, 2006; Pellegrini, 1983). Findings suggested that previous research has largely been unsuccessful in training class inclusion responding to young children (i.e., children below the age of 6). Furthermore, of the studies that showed some training success, these studies either (a) failed to assess generalisation or maintenance of class inclusion responding, or (b) showed disappointing maintenance or generalisation results (McCabe & Siegel, 1987; Pellegrini, 1983). However, the findings of Study 6 indicate that the successful facilitation of class inclusion responding with children even as young as 3 years 5 months is possible, as significant and rapid improvements of correct class inclusion responding were observed for all participants as well as maintenance and generalisation.

These results suggest the potential efficacy of RFT-based procedures to successfully establish class inclusion reasoning within typically developing 3- to 4-year-old children. The findings indicate that the combined application of non-arbitrary training, contextual cues and reinforcement contingencies successfully facilitated class inclusion responding within pre-operational populations. The potential efficacy of non-arbitrary training, in this case providing guidance by showing stimuli in relevant non-arbitrary containment relations, is in accordance with previous RFT work which has shown the efficacy of using such relational training (e.g., Berens & Hayes, 2007).

The success of the training employed in Study 6 prompts additional investigation regarding the effectiveness of the included training procedures with further populations, including younger typically developing children as well as perhaps populations with developmental delay including for example children with autism, Down syndrome or other intellectual disability (Baron-Cohen, Joliffe, Mortimore & Robertson, 1997; Herold, Tényi, Lénárd, & Trixler, 2002; Shulman, Yirmiya, & Greenbaum, 1995). The aim of Study 7 was to examine the efficacy of the procedures used in children with ASD.

Study 7[1]

Given the correlation between categorisation and language ability, it is reasonable to conclude that categorisation skills may be deficient within populations that also evidence deficits in language abilities. This is reflected in the extant research, which has revealed that verbal IQ scores predict category performance in individuals with ASD (Alderson-Day & McGonigle-Chalmers, 2011). According to the current diagnostic and statistical manual of mental disorders, fifth edition (DSM-5), the diagnostic criteria for ASD state that individuals often demonstrate an unusual fixation with parts of objects. Such a bias has been theorised to have significant consequences for categorisation abilities, most notably when faced with the categorisation of stimuli which vary in physical features, but share common global stimulus properties (Fiebelkorn, Foxe, McCourt, Dumas & Molholm, 2013).

ASD is typified by impairments in a number of domains, including cognitive, social and behavioural abilities. For instance, individuals with ASD

[1] This study was conducted in collaboration with Siri Ming, ABA PhD Candidate at NUI Galway.

show evidence of atypical perceptual processing (Dakin & Frith, 2005), in addition to demonstrating difficulty in perceptual learning (Church et al., 2010; Klinger & Dawson, 2001). As such, it is hypothesised that such perceptual deficits mirror irregularities in the cortical structure of individuals with ASD (Markram & Markram, 2010; Rubenstein & Merzenich, 2003; Spencer, O'Brien, Riggs, Braddick, Atkinson & Wattam-Bell, 2000). When one considers the differences which exist between typically developing children and those with ASD in relation to perceptual abilities and learning processes, it is logical to suspect that repertoires of categorisation and category learning may be deficient within ASD populations.

To date, few studies have examined categorisation skills in ASD populations, however, there is a burgeoning research base examining this topic. Thus far, this research has yielded information regarding ASD populations evidencing differences in attention to perceptual similarities between stimuli (Church et al., 2010; Church et al., 2015; Edwards, Perlman & Reed, 2012; Mercado et al., 2015; Perreault, Habak, Lepore, Mottron, & Bertone, 2015; Plaisted, O'Riordan & Baron-Cohen, 1998; Ropar & Peebles, 2008; Tager-Flusberg, 1985), and differences in the generation of prototype (Church et al., 2010; Church et al., 2015; Gastgeb, Dundas, Minshew & Strauss, 2012; Gastgeb, Rump, Best, Minshew & Strauss, 2009; Gastgeb, Wilkinson, Minshew & Strauss, 2011; Klinger & Dawson, 2001; Mercado et al., 2015). Furthermore, research has also indicated a delay in the development of this skill within this population (Johnson & Rakison, 2006), which extends into adolescence and adulthood (Naigles, Kelley, Troyb & Fein, 2013; Gastgeb et al., 2012). Categorisation skills within this population are also characterised by atypical categorisation generalisation (Church et al., 2010; Church et al., 2015). As such, applied behaviour analysis has typically been concerned with augmenting or facilitating categorisation relevant repertoires among individuals with ASD.

Facilitating categorisation repertoires among autistic populations. The field of behaviour analysis has provided important contributions relating to the assessment and facilitation of categorisation skills within developmentally disabled populations. Most notably, in recent years one avenue of research has been concerned with employing derived equivalence paradigms to facilitate categorisation responses (Kobari-Wright & Miguel, 2014; Miguel & Kobari-Wright, 2013; Miguel, Petursdottir & Carr, 2005; Miguel, Petursdottir, Carr & Michael, 2008; Petursdottir, Carr, Lechago & Almason, 2008). This vein of research has incorporated both neuro-typical (NT) young children, in addition to children with a diagnosis of ASD into the training process. The initial stages of training involved providing participants with tact and/or receptive identification training of categorically related stimuli. The stimuli were tacted or identified based upon both their individual (e.g., hammer, drill) and categorical label (tools). Following completion of this phase of testing, participants were then assessed for the presence of emergent categorisation-type skills, which were not explicitly trained within the initial phase of testing.

While this work marks an important step forward in the training of categorisation skills, which have been described as fundamental to cognition and concept learning (Lakoff, 1987), it fails to conceptualise hierarchy in its more complex and arbitrary form, instead focusing on the non-arbitrary manifestation of categorisation abilities. By conceptualising categorisation at a basic level, there is an inherent failure to incorporate a crucial feature of categorisation, which is that categories can be hierarchically related in complex and elaborate infrastructures. For instance, a “fruit fly” is a member of the category “insect”, while “insect” is a member of the category “invertebrate”, and “invertebrate” is a member of the category “animal”. However, not all animals are fruit flies, insects, or invertebrates. In order to be considered truly capable of exhibiting categorisation skills, it is important that individuals respond appropriately in accordance with this feature, which is commonly tested using the class inclusion task. Focusing on the facilitation of the early foundation of categorisation abilities, in the form of class inclusion responding, in individuals with ASD may perhaps generate a number of additional positive outcomes, including improvement in everyday functioning as well as facilitation of acquisition of other intellectual or cognitive skills.

Method

Participants. Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland, Galway, Research Ethics Committee. Consent for conducting the study was obtained from the principal in each respective school. Parental consent was obtained for each child who participated and verbal assent was also obtained from each of the participants. Three individuals with autism participated in this study, which was a non-concurrent multiple baseline design in which participants were pre-assigned to baseline lengths of 3, 5, or 7 sessions. Pt 7.1 was aged 8 years 1 month at the time of testing and had an age-equivalence score of 7 years and 2 months as measured by the Peabody Picture Vocabulary Test, 4th edition (PPVT-4). Pt. 7.2

was 19 years and 1 month of age (PPVT-4 AE = 7yrs 11mths). Pt 7.3, was 9 years and 7 months of age at the onset of the study (PPVT-4 AE = 6yrs and 5mths).

Setting. All participants were enrolled in specialised schools catering for individualised and small group instruction for children with ASD. Pt 7.1 was enrolled in one such school in Bangalore, India. Pt. 7.2 was located in Phoenix, USA, and Pt 7.3 was located in Baltimore, USA. All sessions were performed in a room, separate from the individuals' classroom, at the school in which each participant was enrolled. The sessions were conducted by teachers and behaviour analysts at these schools.

Experimental Design. The study employed a non-concurrent multiple baseline design in which participants were pre-assigned to one of three baseline lengths (3, 5, or 7 sessions). The primary advantage of a non-concurrent multiple baseline design is that it is particularly practical when the research requires participants who have specific behavioural repertoires that are difficult to find within that population. The non-concurrent multiple baseline design is also useful when more rigorous designs are unavailable due to practicalities, resources or time constraints (Gast et al., 2014). For example, the non-concurrent multiple baseline design was employed based upon practical considerations relating to the identification of suitable participants with the necessary skills (determined by screening) to take part in the current intervention. Furthermore, the three participants were ultimately located in three different geographic locations, making the non-concurrent multiple baseline design the most practical experimental design to employ. Participants were pre-assigned randomly to different baseline lengths, as advised by Watson & Workman (1981) and Christ

(2007) to improve the internal validity of the non-concurrent multiple baseline design.

Materials. Class inclusion materials (baseline, training, maintenance and generalisation) were the same as those outlined for Study 6. Tokens served as generalised conditioned reinforcers, which participants could exchange on a Fixed Ratio (FR) 4 schedule for a choice of individualised reinforcers. Praise contingent on correct responses was also provided.

Inter-Observer Agreement (IOA) and Procedural Fidelity

Sessions were observed in person or via video for procedural fidelity for approximately 20% of testing sessions for each of the participant. Due to a lack of resources and issues with access to personnel, only 20% of these sessions were assessed for IOA and procedural fidelity. This constitutes a significant issue regarding the reliability of the results, however, it also reflects the practical issues associated with applied research and the difficulties associated with accessing such populations. Despite this, the results indicated a high level of procedural fidelity throughout the sessions observed, with results ranging from 87.5% to 100% fidelity (average 97%). A second observer collected IOA for 20% of sessions for each participant. The results also indicated a high level of IOA, with 100% agreement in all sessions.

Screening. Each participant was screened for their suitability for the current study prior to their participation. Participants were tested via table top discrete trial procedures to assess their ability to (1) tact all relevant stimuli, (2) correctly respond to ‘yes’ and ‘no’ stimulus identification questions (e.g., “Is this an apple?”), (3) tact the categories of the stimuli within the study, (4) tact quantities of the stimuli from one to ten, and (5) correctly respond to non-arbitrary comparison relation questions (e.g., “Are there more horses or more sheep?”).

Procedure

All baseline, intervention, and generalisation procedures were conducted using the same procedure previously outlined for Study 6. However, the maintenance procedure was slightly modified; maintenance tests for Study 7 involved two tests: one of the trained category (animals) and one of the untrained categories. In this way, any failure to maintain the trained skill could be separated from a failure to maintain the generalised skill. The criterion for maintenance and generalisation of class inclusion responding for this study was 100% correct responses for class inclusion tasks. As individuals with autism typically demonstrate difficulties with generalising and maintaining responding (Paul, 2008), this criterion was selected in order to identify any issues with generalisation and maintenance with the participants and therefore reintroduce them to the intervention so that participants could maximise the benefits of the current intervention. The timing for maintenance assessment varied per participant as a result of participant and experimenter availability. Pts 7.1 and 7.3 were assessed for maintenance of training effects 6 weeks following training cessation, while Pt 7.2 was assessed 8 weeks following the end of training.

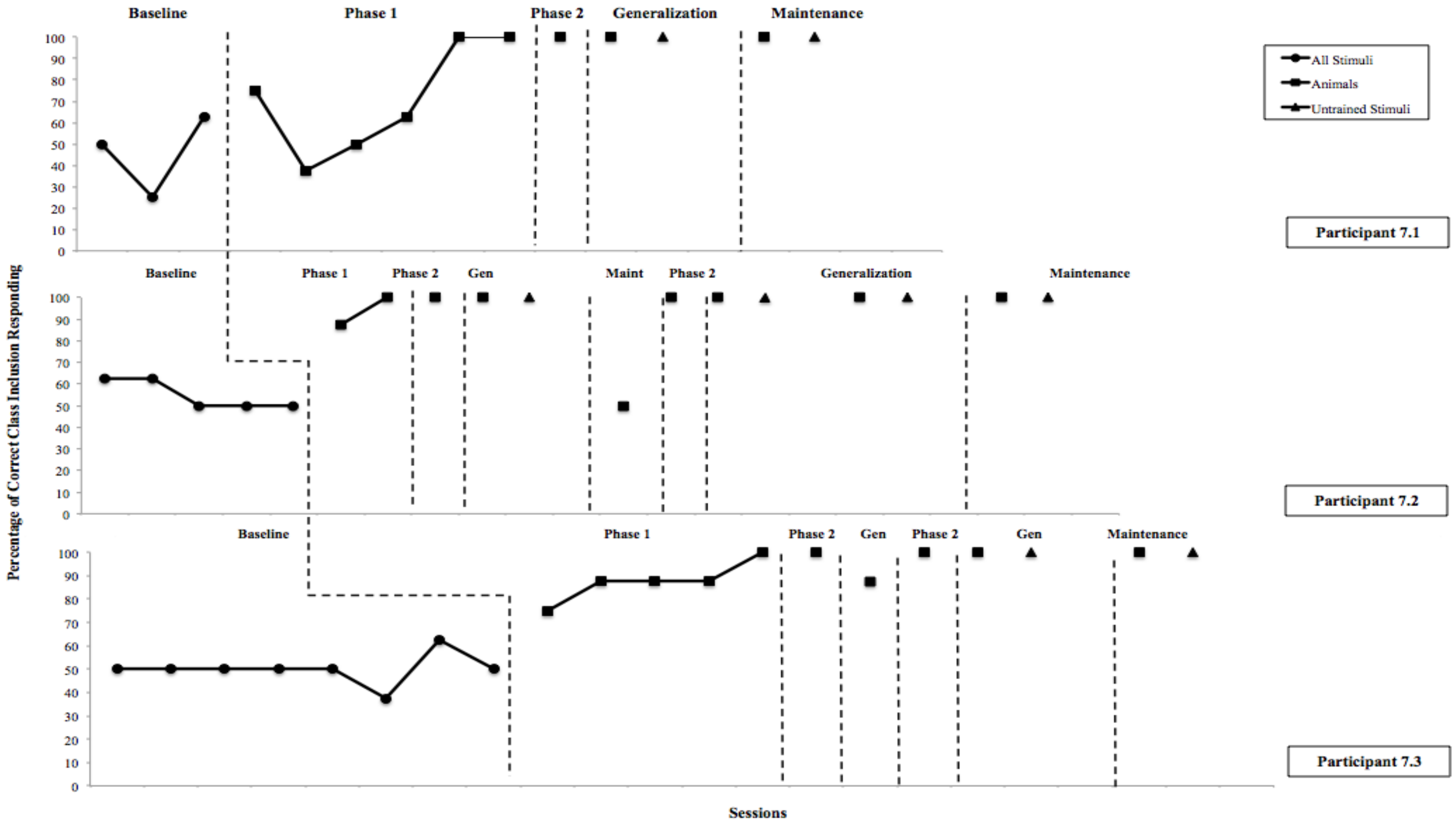
Results and Discussion

Results for Pts. 7.1, 7.2 and 7.3 are displayed in Figure 6.3. During baseline assessments, all three participants showed chance level responding. In order to demonstrate class inclusion responding, an individual must correctly answer all task types. As such, while a participant may have scored 50% correct on class inclusion tasks, this is not considered to be a true display of a class inclusion repertoire. Interestingly, however, within this, there were differences in terms of the patterns seen. Pt. 7.1 responded in a random pattern whereas Pts. 7.2 and 7.3 demonstrated a specific pattern of responding wherein they consistently responded incorrectly to questions addressing the subset with a greater quantity of stimuli (e.g., if there were 5 dogs and 2 cats, any questions about dogs would be incorrect), while questions which addressed the subset containing fewer stimuli were generally answered correctly (e.g., in the previous example provided, cats might be answered correctly). As such, both Pt. 7.2 and Pt. 7.3 showed a reasonably stable baseline score of 50% correct responding, with little deviation from this score. Pt. 7.3 demonstrated an increase in class inclusion responding, to 62.5% in the seventh baseline session, and as such an additional baseline session was performed in order to confirm that an upward trend in class inclusion responding had not begun. The results of this additional session indicated that he had returned to the previous level of 50% correct.

Following baseline assessment, Pt. 7.1 was introduced to the initial phase of intervention (Phase 1) and reached mastery criterion for this level after five sessions. However, as a result of experimenter error, Pt. 7.1 was exposed to an additional session of Phase 1, the results of which indicated that Pt. 7.1 had again demonstrated 100% class inclusion responding. When introduced to the second phase of intervention, Pt. 7.1 met mastery criterion after one session. Pt. 7.1 was then assessed for generalisation of trained and untrained categories and was found to meet the criterion for generalisation (i.e., 100%). Six weeks following the final post-intervention probe, Pt. 7.1 was tested for maintenance of class inclusion responding and demonstrated maintenance of class inclusion responding (100%) for both trained and untrained categories.

Pt. 7.2 responded to Phase 1 of class inclusion training immediately, as following an initial error made during the first trial presented, he then provided correct class inclusion responses for the remainder of the training session, and met mastery criterion for Phase 1 during the second session of testing. Once exposed to Phase 2 of training, Pt. 7.2 demonstrated instant correct responding within the first session. When tested for generalisation of class inclusion responses to both trained and untrained categories, Pt. 7.2 responded correctly to all trials with 100% correct responding for both generalisation tests.

Figure 6.3. Correct class inclusion responses of individuals with ASD.



Note. P 1 = Phase 1; P 2 = Phase 2; Gen = Generalisation; Maint = Maintenance

Due to a combination of factors, including holidays, illness and the availability of staff, Pt. 7.2 was tested for maintenance of class inclusion responding 8 weeks after the final post-intervention probe. The results of this test indicated that his class inclusion responding had returned to that of baseline levels. He was then re-introduced to Phase 2 training, during which he demonstrated correct responding to all class inclusion trials within this first session. Pt. 7.2 was again assessed for generalisation of class inclusion responding for both trained and untrained stimuli, immediately after this training session, and in an additional assessment session two days later. Two weeks following this assessment, Pt. 7.2 was then tested for maintenance of class inclusion responding. During this session, Pt. 7.2 made one error within the test assessing maintenance of the trained category, but successfully responded to all class inclusion trials for the untrained category. An additional maintenance test was conducted six weeks following training, which indicated that Pt. 7.2 had successfully maintained class inclusion responding at 100% levels for both trained and untrained categories.

Following baseline assessment, Pt. 7.3 met mastery criterion for the initial phase of training after five sessions. When introduced to the second phase of testing, Pt. 7.3 instantly demonstrated correct class inclusion responding, of 100% within the first session. However, when introduced to the first post-intervention probe with trained stimuli, he made a number of errors. As such, he was reintroduced to the second phase of training, in which his performance increased to 100% levels. When re-exposed to the post-intervention probes for generalisation of class inclusion responses to both trained and untrained stimuli, Pt. 7.3 met generalisation criterion for both tests. In addition to this, when tested for maintenance of class inclusion responding six weeks following the final post-intervention probe, Pt. 7.3 also met mastery criterion (i.e., 100% correct) for both trained and untrained stimulus sets.

In summary, all participants were successfully trained to perform correct class inclusion responses and also showed both generalisation and maintenance across a number of stimulus sets. This study is the first to successfully implement a training procedure designed to facilitate class inclusion responding with individuals with a diagnosis of ASD. In view of the lack of research within this area, the current study offers valuable information concerning effective procedures to successfully establish class inclusion responding within autistic populations. It further contributes to the extant literature on class inclusion responding.

This study also included some limitations however which needed to be addressed. For instance, the study employed a non-concurrent multiple baseline across participants design in which participants were randomly pre-assigned to a baseline length. Although random pre-assignment was used in order to improve the internal validity of the design (see Christ, 2007; Watson & Workman, 1981), in the case of Study 7 stable baseline responding was not observed for two participants (i.e., Pt. 7.1 and Pt. 7.3) prior to the predetermined introduction of an intervention and as such constitutes a methodological limitation of this study. In addition to this, it was difficult to assess participants' generalisation of class inclusion responding to untrained categories as insufficient information had been gathered regarding the stimulus categories used in baseline. As such, the generalisation results within Study 7 must be interpreted with caution based upon this methodological limitation. Based on the results of Study 7 it is also unclear as to what phase of training facilitated class inclusion responding. For instance, it is possible that the participants could have successfully established class inclusion responding with Phase 2 training only.

Study 8[2]

Study 7 showed successful training and testing of class inclusion in individuals with ASD. Nevertheless, as discussed, it featured several limitations including participants' unstable responding during baseline assessment, the use of a non-concurrent baseline design, and incomplete information regarding the stimulus categories used during baseline. In the current study (i.e., Study 8) these were addressed by employing a concurrent multiple baseline across participants design in which participants were only introduced to training if (a) they demonstrated stable baseline responding, and/or (b) the previous participant had successfully completed training and met generalisation criterion (i.e., 100% across all categories). Following from this point, all stimulus categories employed during baseline assessments were recorded in order to provide a more accurate gauge of generalisation. Study 8 also only employed Phase 2 of training, as it was hypothesised that Phase 1 (employed in Studies 6 and 7) may not have been necessary and resulted in an unnecessarily long training procedure. Finally, as two of the three participants in Study 7 demonstrated difficulties with either generalisation or maintenance of class inclusion responding, it was decided to alter the mastery criterion for class inclusion responding. This was changed to 100% correct class inclusion responding across two consecutive sessions.

---

[2] This study was conducted in collaboration with Patrycja Zagrabska, ABA MSc Candidate at NUI Galway.

Method

Participants. Prior ethical approval for recruitment of participants for this study was obtained from the National University of Ireland, Galway, Research Ethics Committee. Consent for conducting the study was obtained from the principal of the school. Parental consent was obtained for each child who participated and verbal assent was also obtained from each of the participants. Four individuals with a diagnosis of ASD participated in the current study, which was a concurrent multiple baseline design. Pt. 8.1 was 16 years old and was diagnosed with a co-occurring mild general learning disability (PPVT-4 Age Equivalent score = 9yrs 5mths). Pt. 8.2 was 15 years old with a co-occurring mild intellectual disability (PPVT-4 AE = 7yrs 1 mth). Pt. 8.3 was 13 years 7 months old with a mild general learning difficulty, gelastic epilepsy (PPVT-4 AE = 10yrs 2mths). Pt. 8.4 was 12 years 2 months and was diagnosed with a co-occurring mild general learning disability (PPVT-4 AE = 8yrs 1mth).

Setting. All participants were students enrolled in a special school dedicated to the education of children and teenagers with moderate, severe and profound developmental disabilities and ASD in the West of Ireland. All sessions were performed in a room, separate from the individuals' classroom, and the sessions were conducted by a behaviour analyst within the school.

Experimental Design. A modification of the previous research included a concurrent multiple baseline design, rather than the previously employed non-concurrent multiple baseline design. Within a concurrent multiple baseline design, baselines are established simultaneously, while all data are also collected at the same time. All participants were introduced to baseline conditions until a stable baseline was observed among at least one participant. This participant was then introduced to class inclusion training while the remaining participants remained in baseline conditions. Participants were subsequently introduced to class inclusion training if: (a) the previous participant had successfully passed the class inclusion training phase and demonstrated generalisation, and (b) the participant had a stable baseline level of class inclusion responding (See chapter 4 for information regarding the benefits of this experimental design).

Screening and Materials. Each participant was screened for their suitability for the current study using the same screening assessments and materials as outlined in Studies 6 and 7. In a further modification to the previous studies, all participants were assessed for their preferred stimuli using a questionnaire provided to caregivers and teachers. Based upon the results of this assessment, participants were told that they could choose what activity or object they were working for prior to the commencement of the session. All participants also received positive social reinforcement contingent upon correct responses.

Inter-Observer Agreement (IOA) and Procedural Fidelity

Sessions were observed in person for procedural fidelity for approximately 20% of testing sessions for each of the participant. As with Study 7, the collection of IOA and procedural fidelity data was hampered due to a lack of resources and limited access to the participants, thereby negatively impacting the reliability of the results generated within the current study. The results indicated a high level of procedural fidelity throughout the sessions observed, with 100% procedural fidelity recorded across all sessions. IOA was also recorded for 20% of sessions for all participants. The results indicated a high level of IOA, ranging from 90 – 100% (*M* = 95%) agreement in all sessions.

Procedure

Baseline and Class Inclusion Training. The baseline condition was identical to that of baseline conditions within the previous studies; however, the current study also recorded the category stimuli used within each baseline trial and the participants' response to the class inclusion question. In a further modification to Studies 6 and 7, only Phase 2 of the previously described procedure was employed within the current study with a provision that should participants fail on this phase, then the equivalent of Phase 1 from the previous studies would be introduced at that point. Mastery criterion was also changed within the current study, such that a participant had to demonstrate 100% correct responding across two successive sessions in order to achieve mastery criterion and progress to the generalisation phase of the study. This mastery criterion was selected based upon the results of Study 7. Within Study 7, two out of three participants demonstrated difficulties with the generalisation and maintenance of responding, however, this was addressed by exposure to one additional training session. As such, Study 8 introduced a more stringent mastery criterion in order to ensure generalisation and maintenance results for participants.

Generalisation and Maintenance. In an additional modification to the previous studies, once participants had successfully completed training, they were then assessed for their generalisation of class inclusion responding using the trained stimulus set (i.e., animals) without the presence of the boxes. If participants then demonstrated generalisation with this stimulus set, they were subsequently tested for generalised class inclusion responding with untrained stimulus sets (i.e., vehicles, clothes and fruit) across all 8 class inclusion questions, totalling 24 questions in all as opposed to the original 8 questions asked within Studies 6 and 7. Maintenance tests were identical to those within Studies 6 and 7 (i.e., 8 class inclusion questions were used to assess maintenance across four categories), but were conducted within a shorter period post-training (i.e., 2 weeks) due to time constraints. The criterion for maintenance and generalisation for this study was identical to that of Study 7 (i.e., 100%, see Study 7 for rationale).

Results & Discussion

The results of Pt. 8.1, Pt. 8.2, Pt. 8.3, and Pt. 8.4 are presented in Figure 6.4. As previously outlined, the category stimuli used within each baseline trial was also recorded for each participant in addition to correct and incorrect responding. These data are displayed in Table 6.2.

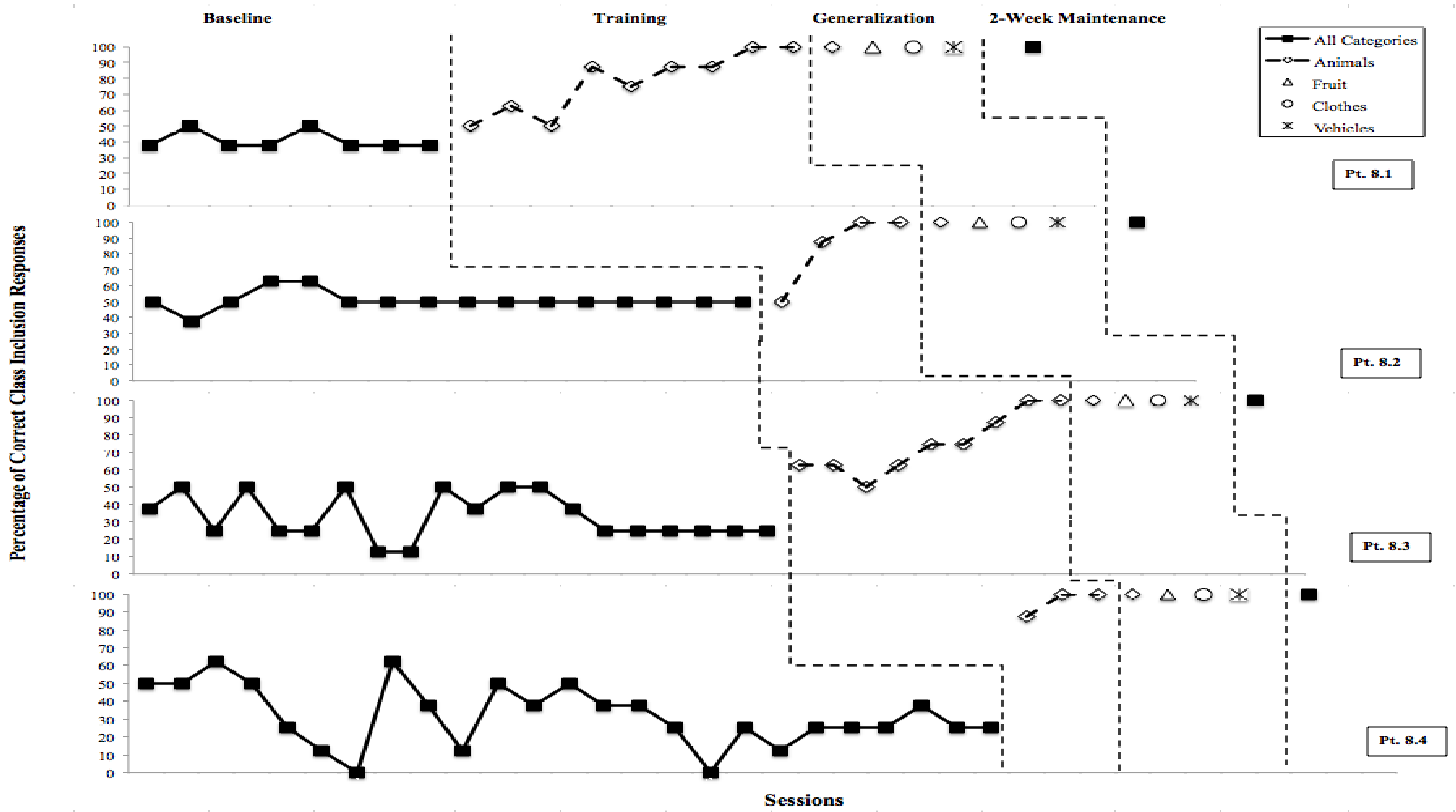
Table 6.2. Categories assessed and number of correct responses per participant.

Category	**Participants Correct Class Inclusion Responses**			
	Pt. 8.1	***Pt. 8.2***	***Pt. 8.3***	***Pt. 8.4***
Animals	6/16 (38%)	19/32 (59%)	9/40 (23%)	16/50 (32%)
Fruits	6/16 (38%)	12/32 (38%)	16/40 (40%)	15/50 (30%)
Clothes	9/16 (56%)	22/32 (69%)	13/40 (33%)	14/50 (28%)
Vehicles	5/16 (31%)	12/32 (38%)	15/40 (38%)	19/50 (38%)
Total	26/64 (41%)	65/128 (51%)	53/160 (33%)	64/200 (32%)

Baseline scores for all participants remained stable prior to the introduction of the training intervention. The sequential introduction of class inclusion training for each participant resulted in considerable increases over baseline levels. These changes were obtained without corresponding changes in the baselines of the remaining untreated participants thus showing the causal role of the intervention. Results for generalisation and maintenance are also presented in Figure 6.4. These data show both generalisation and maintenance across participants.

During baseline conditions, Pt. 8.1 was the first participant to demonstrate a stable rate of baseline class inclusion responding, between 37.5% and 50% correct (M = 40.63%). Pt. 8.1 demonstrated an increase in class inclusion responding once introduced to the training procedure and reached mastery criterion after a total of nine training sessions. Following the cessation of training, Pt. 8.1 was then tested for generalisation of class inclusion responding across the trained stimulus set (i.e., animals) and untrained stimulus sets (i.e., fruit, clothes and vehicles). The results of these assessments were that Pt. 8.1 showed generalisation of class inclusion to both trained and untrained stimulus sets. Furthermore, Pt. 8.1 also demonstrated maintenance of class inclusion responding at two weeks post training.

Figure 6.4. Correct class inclusion responses of individuals with dual diagnoses.



Once Pt. 8.1 had successfully demonstrated generalisation of class inclusion responding, the next participant with a stable baseline of class inclusion responding was then introduced to training. Pt. 8.2 demonstrated a similarly stable rate of baseline class inclusion responding, between 37.5% and 62.5% correct (M = 50.78%). Once introduced to class inclusion training, Pt. 8.2 met mastery criterion after four training sessions and subsequently demonstrated generalisation to both trained and untrained stimulus sets. Pt. 8.2 also demonstrated successful maintenance of class inclusion responding two weeks following the cessation of training.

Following Pt. 8.2's successful acquisition and generalisation of class inclusion responding, Pt. 8.3 was subsequently introduced to training as he had a more stable rate of baseline responding than that of the remaining participant, Pt. 8.4. Upon visual analysis, Pt. 8.3's baseline responding was more variable than that of the previous participants, but was relatively stable with correct responses between 12.5% and 50% (*M* = 33.13%). Pt. 8.3 met mastery criterion for this phase of training after nine training sessions and also successfully demonstrated generalisation across trained and untrained stimulus sets.

Pt. 8.4 was the final participant introduced to class inclusion training. Prior to training this participant demonstrated a variable and instable rate of baseline responding, demonstrating correct class inclusion responding between 0 and 62.5% (M = 32%). Once Pt. 8.4 was introduced to class inclusion training, he demonstrated successful acquisition of this repertoire following three training sessions. Pt. 8.4 also demonstrated successful generalisation of class inclusion responses across both trained and untrained stimulus sets.

This study once again showed successful training of class inclusion in participants with ASD who showed deficits in this respect beforehand. This successful training used a similar (though better controlled) training protocol to that employed in Study 7. In addition, this study showed this successful training with individuals each of whom had not alone ASD but additional diagnoses also including intellectual disabilities and learning disabilities, indicating once again the potential flexibility and utility of this paradigm.

Study 8 extended from Study 7 by modifying the existing training procedure in several ways including using a concurrent multiple baseline design, recording the stimulus categories used during baseline, using only Phase 2 of training and altering the mastery criterion for class inclusion responding. By employing a concurrent multiple baseline design, rather than the non-concurrent multiple baseline designs previously employed, a researcher can reasonably assume that the changes observed in class inclusion responding are as a result of intervention, rather than extraneous factors within the participants' environment. Furthermore, the current study also demonstrated generalisation and maintenance of treatment effects across a number of stimulus sets. The modification of the generalisation assessment to test all four stimulus sets with all eight class inclusion questions (totalling 40 class inclusion questions) demonstrated a more reliable measurement of generalisation and was introduced to determine true generalisation across all stimulus sets. In view of the scarcity of literature and research in this area, this study presents practitioners with valuable information concerning effective training procedures to establish this repertoire in individuals with complex diagnoses, and also contributes to the limited volume of research relating to the facilitation of class inclusion responding more generally.

General Discussion

As outlined in the introduction to the current chapter, passing the “class inclusion” test as described here is considered to be an important developmental milestone within mainstream (and more specifically cognitive developmental) psychology. For example, Piagetian psychologists regard this skill as a significant indication of cognitive development; specifically, they see it as indicative of a child’s reaching the “concrete operational” stage of development.

Behaviour analysts including RFT researchers do not subscribe to a stage based account of development and therefore would not see passing the class inclusion test as indicating that an individual has reached a particular stage of development. Instead they would see passing class inclusion as one repertoire within a more global repertoire of hierarchical classification which is itself seen as a very important aspect of negotiating the human world. In this sense passing a class inclusion test is still indicative of something - namely, how advanced a child’s classification repertoire is. More specifically, passing class inclusion suggests ‘understanding’ with respect to the relation of members to classes and classes to each other. Of course the term ‘understanding’ is a nebulous one and something that might traditionally be more associated with a cognitive or mentalistic than a behaviour analytic approach. The advance made by RFT of course is that RFT describes such understanding in technical (functional analytic) terms as patterns of relational responding. In the case of class inclusion, the patterns of relational responding include both arbitrarily applicable hierarchical relations (stimuli being ‘members’ of classes while classes include those stimuli; classes being members of other classes) as well as non-arbitrary more/less relations. It is presumably only when a child has acquired these repertoires to at least some extent that they will be able to successful pass class inclusion tests. Non-arbitrary more/less relations would be acquired much earlier than arbitrary hierarchical relations and thus it is the child's capacity with the latter that would be probed by class inclusion.

In fact, however, in one sense class inclusion may be even slightly more advanced a skill than even this. In other words, it might go beyond even being able to respond based on hierarchical relations alone. More specifically, it might require responding in accordance with both hierarchical and more than/less than relations at the same time – in other words, to show a type of combinatorial derivation involving more than one type of relation. This skill is something that requires that the individual has sufficient training or exposure to hierarchical relations in particular and thus this is why this task is indicative of this relational repertoire. Of course training the capacity to correctly answer class inclusion tasks will strengthen this repertoire further and this will likely boost the child's ability to understand and problem solve in their environment.

The capacity to respond correctly in accordance with hierarchical relations is an important one. Often applied behaviour analysis has addressed the facilitation of categorisation skills in developmentally delayed populations by focusing on linking the names of varying stimuli with the labels of the various categories the stimuli belong to. For example, a child might be taught to tact a dog as both a dog and as an animal. However, the argument can be made that if individuals fail to comprehend the relations between these arbitrary category labels, then they might simply be learning little more than to respond with more than one name or label for the same stimulus, or at best, to respond with different labels under contextual control. If they are not learning particular to respond in accordance with particular patterns of generalised relations between stimuli then their understanding is limited. In more conventional language we might say that they fail to truly comprehend categories and categorisation as a whole.

This is why an RFT approach to training can help in teaching both typically developing children as well as populations such as those with ASD or otherwise that have particular deficits in their verbal / cognitive repertoire. A key purpose of this thesis as a whole has of course been to focus on understanding the establishment (either through exposure to naturalistic contingencies in the environment or through direct targeted training) of relational responding involved in classification. By gaining an understanding of the contingencies involved, what is learned when testing and training typically developing children can then be applied also with children and individuals with deficits in their relational repertoire. In this thesis containment and hierarchical relations were extensively tested and trained in groups of typically developing children, providing insight into typical developmental acquisition of containment and hierarchy relations, their correlations with linguistic and cognitive development overall as well specifically with classification as a repertoire as well as information regarding the capacity to train children using relational framing techniques and some indication of the potential of such training. Due to limitations of time and resources this same assessment and training could not be provided to individuals with ASD and thus this is one area for expansion in the future. However, as reported in the current chapter, this thesis has gone on to explore one potentially important repertoire of classification using an RFT approach, namely class inclusion, and with regard to this particular skill, the possibility of extension of the RFT paradigm to children with ASD has been explored and the utility of the paradigm has been confirmed. The concept of MET of a relational pattern under precise and consistent contextual control and utilising non-arbitrary guidance to facilitate the establishing of a more arbitrary pattern are all key hallmarks of an RFT approach and have allowed training, maintenance and generalisation of the derived relational pattern involved with very young typically developing children as well as with children with ASD.

Of course much further work remains to be done in this case. Many further questions remain to be answered. For example, this study considered generalisation of the trained repertoire, but only in the relatively narrow sense of generalisation to alternative stimulus sets. Studies in previous chapters considered generalisation in a broader sense by considering performance in alternative tests of classification as well as language more broadly. Perhaps consideration of generalisation in some broader sense such as this might be helpful in this case also. Might training in class inclusion influence performance on alternative tasks or measures of classification, such as the Children's Category Test (CCT; Boll, 1993) for example? Furthermore, while it is theorised that class inclusion responding forms a cornerstone of hierarchical responding, a further avenue of research could determine the extent to which class inclusion training may facilitate such hierarchical responding repertoires. The inverse question to this of course is whether sufficiently extensive training in containment and hierarchical relational repertoires themselves might make class inclusion responding more likely. Perhaps training in key elements of that repertoire (e.g., contextual control over hierarchical relations) without focusing on the class inclusion paradigm itself might nevertheless make successful performance on this repertoire more likely.

With regard to the population of ASD individuals that were employed in the current studies, in both Studies 7 and 8, individuals with age equivalent PPVT-4 scores of around 7 were employed as participants. Of course, as per screening and baseline testing, none of these individuals could show class inclusion responding and had to be trained to do so. However, as these individuals were ostensibly performing, at least according to the PPVT-4, at the verbal-cognitive level of a seven year old, then according to both previous theory and empirical research on class inclusion perhaps their overall repertoire was such as to facilitate relatively ready learning of class inclusion. It might be interesting to future to work with individuals with a PPVT-4 age equivalence score significantly below 7 years to see whether they might learn class inclusion based on a protocol such as the current one quite as readily.

Additional avenues of research could also be concerned with analysing the efficacy or effectiveness of various components of the protocol. For instance, the necessity for the incorporation of boxes, designed to highlight the non-arbitrary containment relationship between the testing stimuli, might be assessed to determine the importance of this component. Future work might systematically manipulate the presence of this component to assess its importance. Such work might suggest that the boxes might not be necessary components of the protocol for all populations. In that case it might be interesting to examine for which populations they are necessary and how effective they are in those cases.

In conclusion, the studies reported in this chapter constitute an advance in the domain of class inclusion research as well as with respect to young children and individuals with developmental delay. Nevertheless it is preliminary research that prompts further investigation in this area within the context of continuing work on relational framing repertoires involved in classification more broadly.

Chapter 7

General Discussion

The aim of the research reported in this thesis was to assess and train relational repertoires relevant to hierarchical classification in a range of individuals, including primarily young children but also including individuals diagnosed with Autism Spectrum Disorder (ASD) who showed relevant deficits. As described in the opening chapter, the majority of previous research addressing hierarchical classification has originated from the cognitive developmental field. This work has, for example, identified and provided data regarding properties of hierarchical classification, including transitive and asymmetric class containment and unilateral property induction, in both child and adult populations. Meantime, little or no research into this repertoire had come from behaviour analysis. In the last few years, however, especially with the advent of Relational Frame Theory (RFT), this has begun to change. A number of studies have begun to explore this repertoire as particular patterns of relational framing including especially containment and hierarchical framing. However, most if not all of this latter work has been done with adults with already well-established verbal repertoires. As such the aim of the current doctoral research was to extend this research by investigating relational framing relevant to hierarchical classification in populations, including in particular young children, in whom this repertoire is not already strongly established.

The current research programme started (Study 1) by assessing hierarchical and containment relational repertoires in young typically developing children ranging in age and correlating framing performance with that on standardised measures of cognition, language and categorisation. The studies that followed (Studies 2-8) investigated the efficacy of multiple exemplar training (MET) for establishing containment (2 & 3) and hierarchical (4 & 5) relational framing as well as another repertoire relevant to hierarchical classification, namely class inclusion (6-8) in both typically developing children as well as individuals with ASD. In the present ‘General Discussion’ chapter I’ll first provide a brief overview of these studies. Following this, I’ll discuss theoretical implications and limitations of this work, as well as potential avenues for future research.

Overview of PhD Studies

Study 1 (Chapter 3)

As explained in Chapter 2, RFT sees arbitrarily applicable derived relational responding or relational framing as critical for language and cognition, and RFT research has shown a strong correlation between performance in accordance with various patterns of relational framing (e.g., coordination, distinction, comparison) and that on standardised measures of language and cognition (Cassidy et al., 2011; 2016; Moran et al., 2014, 2015; O’Toole & Barnes-Holmes, 2009; Parra & Ruiz, 2016; Vizcaíno-Torres et al., 2015). Prior to the current programme, relatively little research in this or other respects had been conducted with regard to containment and hierarchical relational responding. No research had been conducted with children or had attempted to correlate containment or hierarchical framing with language or cognition. Also, while there had been some research into the ages at which other frames such as coordination and distinction might emerge (see Lipkens, Hayes & Hayes, 1993), there was a dearth of information regarding when containment and hierarchy repertoires begin to be acquired.

Accordingly, the aim of Study 1 was to extend prior research by assessing containment and hierarchical relational framing in a relatively large sample of children across a range of ages (i.e., 3 – 8 years old) and correlating performance with that on standardised measures of language (i.e., Peabody Picture Vocabulary Test, 4th edition; PPVT-4), cognition (i.e., Stanford Binet, 5th edition; SB5) and categorisation (i.e., Children's Category Test; CCT; Piagetian Class Inclusion test). Results showed that both containment and hierarchical framing correlated strongly with both language (i.e., PPVT-4) and cognitive (i.e., SB5) performance and moderately with categorisation (i.e., CCT and Piagetian class inclusion). Findings provided information regarding the initial age at which these repertoires tend to be acquired and, by showing correlation of performance with age, indicated that the framing repertoire strengthens gradually over time. Finally, participants were also re-tested 1-2 weeks after initial assessment and results showed high reliability for the framing measures.

These results from Study 1 are in line with those of previous empirical research (e.g., Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; Moran et al., 2014, 2015), which have found that relational framing is positively correlated with performance on standardised measures of language and cognitive performance. They extend the findings of past research by providing information on containment and hierarchical relational framing in particular. They may also offer practical information regarding these framing repertoires for practitioners working with specific populations, including both typically developing children as well as children with developmental delay; more specifically, they might shed light on what might qualify as potential deficits for children in the former group and they might facilitate comparison (i.e., versus typically developing children raised in a relatively conventional environment) for those in the latter group.

Studies 2 and 3 (Chapter 4)

Study 1 had provided data regarding age groups that tended to respond at or below chance levels on measures of arbitrary containment and hierarchical relational responding. Based on this information, Studies 2-5 investigated the potential for training particular repertoires assessed in that initial study to age-appropriate cohorts. Studies 2 and 3 recruited 5 year old children for assessment and training in arbitrary containment relational responding, while Studies 4 and 5 recruited 6 year old children for assessment and training in arbitrary hierarchical relational responding.

Study 2 was a pilot study, the aim of which was to teach arbitrary containment in a typically developing 5-year-old female using MET in the context of a multiple baseline across components design (i.e., with mutual entailment; transformation of stimulus function with two stimuli; combinatorial entailment; transformation of stimulus function with three stimuli as components). Results showed that the intervention successfully established arbitrary containment relations including both generalisation to novel stimuli and 1-month maintenance. Study 3 sought to extend Study 2 by using MET to train containment relational framing in three participants in a multiple baseline design across both participants and components. In addition, this study assessed the effect of training on performance on measures of language (using the PPVT-4) and categorisation (using the class inclusion and CCT tests) by comparing experimental participants' performance on these measures before and after training with that of a control group. Study 3 showed successful training, generalisation and 6-month maintenance. Furthermore, while both the trainees and controls demonstrated gains in PPVT-4, CCT and class inclusion performance, the trainees all showed comparably greater gains in these repertoires. Given the very small group sizes, no reliable conclusion can be made on this basis but the results are at least suggestive.

In summary, Studies 2 and 3 showed the efficacy of MET for establishing arbitrary containment relational responding in typically developing five-year-old children, including both generalisation and maintenance of training effects. The findings of Study 3 additionally suggest the potential impact of arbitrary containment training on language, categorisation and class inclusion responding.

Studies 4 and 5 (Chapter 5)

As just described, Studies 2 and 3 examined the potential for establishing arbitrary containment relational framing in 5 year olds. Studies 4 and 5 then sought to examine whether arbitrary hierarchical relational framing, an arguably more abstract and complex repertoire, might be established in 6 year olds. Study 4 aimed to teach arbitrary hierarchical framing to a typically developing 6-year-old within a multiple baseline across four relational components, using a design similar to that of Study 2. Study 5 taught arbitrary hierarchy to three typically developing 6-year-olds but this time, analogous to Study 3, a combined multiple baseline design across both (relational) components and participants was employed. Also analogous to Study 3, Study 5 additionally examined the effect of the training on children’s performance in additional tests of language and categorisation.

Study 4 taught arbitrary hierarchy across four components, including mutual entailment, transformation of functions via mutual entailment, combinatorial entailment and transformation of functions via combinatorial entailment, to a typically developing 6-year-old female. This study successfully established arbitrary hierarchical relational responding across all four components within a four-week period and also showed generalisation which maintained for up to one-month post-training. Study 5 expanded on this by employing a combined multiple baseline design across both participants and relational components to establish arbitrary hierarchical relational responding in three typically developing 6-year-olds. Analogous to Study 3, both these three training participants and three control participants were assessed for language (PPVT-4) and categorisation (CCT and class inclusion) performance both before and after training. All three training participants showed arbitrary hierarchical relational responding, which generalised and maintained up to six months post-training, while the controls showed minimal gains in this repertoire. All six children (both training and control group) showed some increase in their raw scores for language (PPVT-4) and categorisation (CCT, class inclusion) in follow up testing but the training group demonstrated greater gains than the controls. As in the case of Study 3, this outcome is suggestive only however, given the very small sample size involved, but the effects certainly warrant further investigation.

In summary, both Studies 4 and 5 showed the efficacy of MET for establishing arbitrary hierarchical relational responding, including generalisation and maintenance, in typically developing six-year-old children, while Study 5 also provided evidence suggesting that such training may positively impact language and categorisation repertoires.

Studies 6 -8 (Chapter 6)

Studies 6-8 focused on training ‘class inclusion’ responding in young children. Class inclusion responding, which involves responding to a group of stimuli simultaneously as both members of a class and also members of a superordinate class that contains that class, is seen within the mainstream cognitive developmental literature as a useful test of hierarchical categorisation responding (Thomas & Horton, 1997). From an RFT perspective, class inclusion is a pattern of relational responding that involves both arbitrary relations (in the abstract relationship between members and classes and between classes and a superordinate class) as well as non-arbitrary relations (given that the class members are usually arrayed before the child so that they can see that there are more of one class than the other). Given the fact that class inclusion responding is seen as an important progress marker towards understanding of hierarchical classification, in the final empirical chapters of the current programme, it was decided to use an RFT approach to assess and train class inclusion responding in young typically developing children and children with developmental disabilities.

Study 6 employed a non-concurrent multiple baseline across participants design to evaluate the effectiveness of a non-arbitrary containment training procedure to facilitate class inclusion responding in three typically developing 3- to 4-year-olds. Training involved the application of a procedure in which responding was guided by highlighting both contextual cues (e.g., ‘the class of animals contains dogs’, etc.) as well as non-arbitrary containment relations (by using transparent boxes to contain classes and sub-classes). All three children successfully completed training and showed generalised responding which was maintained for at least one month afterwards.

Study 7 employed the same procedure as in Study 6 to establish class inclusion responding in three individuals (age range 8 – 19 years) with autism spectrum disorder, once again in the context of a non-concurrent multiple baseline across participants design. All three participants were assessed for class inclusion responding as well as language (PPVT-4 age equivalence scores ranged from 6.5-8). All three were successfully trained. Two out of three showed immediate generalisation to novel stimulus sets, while the third did so following further training. Two out of three showed maintenance six weeks post-training for both trained and untrained stimulus sets. The third initially failed but after further training showed maintenance 6 weeks following that training. These results indicate the feasibility of training class inclusion responding in this population using MET.

Study 8 addressed the limitations of Studies 6 and 7. First, a concurrent rather than non-concurrent multiple baseline design was used. This helped to achieve stable baseline responding prior to training, something which had not been satisfactorily achieved in the two previous studies. Second, during all baseline assessments the stimulus category employed on every trial was noted in order to accurately gauge generalisation to non-trained stimulus sets post-training. Third, participants were only exposed to one phase of class inclusion training (i.e., Phase 2 of the original protocol). Finally, the mastery criterion for class inclusion responding was also adjusted (i.e., to two successive sessions of 100% correct). Participants were four males (12 – 16 years old) with ASD and a co-occurring developmental disorder. As in Studies 6 and 7, all were assessed on intake for class inclusion responding as well as on the PPVT-4 (AE scores = 7 years 1 month - 10 years 2 months). Findings showed that the training procedure successfully facilitated class inclusion, generalisation (in both trained and untrained stimulus sets) and maintenance (for two weeks) in all four participants.

In summary, Studies 6-8 showed the efficacy of MET for establishing class inclusion responding, including generalisation and maintenance, across a range of participants, including both typically developing children and individuals with developmental delay.

Theoretical Implications

Having briefly reviewed the findings of the studies conducted in the current research programme, I will next discuss and explore the implications of these findings with respect to a number of research domains including Relational Frame Theory, behaviour analysis and cognitive developmental psychology.

Implications for Relational Frame Theory Research

The work reported here presents further support for the Relational Frame Theory position that arbitrarily applicable relational responding is the core operant underlying human language and cognition. RFT research had already provided considerable evidence for this claim (see for example, Cassidy, Roche, Colbert, Stewart & Grey, 2016; Cassidy, Roche & Hayes, 2011; Cassidy, Roche & O'Hora, 2010; Dymond & Roche, 2013; Hayes & Stewart, 2016; Hughes & Barnes-Holmes, 2016). The current data further extends this evidence.

More specifically, the current research has correlated a protocol assessing non-arbitrary containment, arbitrary containment and arbitrary hierarchy with cognitive and linguistic performance in typically developing children ranging from 3 to 8 years (Study 1) and shown a strong positive relationship. It has also provided some (albeit relatively more limited) evidence that training in arbitrary containment (Study 3) and arbitrary hierarchy (Study 5) could positively affect intellectual performance. Study 1 was also the first study to examine the relationship between containment and hierarchical relational responding and performance on a standardised measure of categorisation (i.e., the CCT) with the results suggesting that performance on these relational repertoires and on the CCT were indeed correlated ($\rho = .615$; $p < .001$). While this result might have been anticipated by RFT, nevertheless, this was the first study to actually empirically demonstrate a relationship between categorisation and specific (theoretically related) types of relational framing. The results are indicative and further work is needed. However, at the least this goes towards further supporting the RFT thesis that relational framing underlies human language and cognitive repertoires in general and that particular frames may be useful in understanding particular cognitive repertoires.

Apart from just assessing containment and hierarchical relational framing (i.e., relational framing of classification), the current research has also successfully trained these frames in children. This work augments previous studies that have established theoretically important relational framing repertoires in children (Barnes-Holmes et al., 2004; Berens & Hayes, 2007; Hayes & Stewart, 2016) thus adding to collective evidence of the operant (and more specifically in this context, trainable) nature of relational framing. In addition, the current research used multiple baseline designs to do so, thus providing a well-controlled demonstration of the effects of training (Cooper, Heron & Heward, 2007; Horner, Carr, Halle, McGee, Odom & Wolery, 2005). This further bolsters the strength of the evidence and supports use of the particular intervention used in this case.

The current results also showed robust maintenance of the effects of relational frame training. Without follow-up tests, it cannot be argued with any certainty that a genuine and permanent change in relational responding, or other repertoires, might have been produced as a result of training. Nevertheless, maintenance tests are not used as frequently as they should be, either in RFT research or (arguably) otherwise (see Berens & Hayes, 2007; Hayes & Stewart, 2016; Kilroe, Murphy, Barnes-Holmes & Barnes-Holmes, 2014). Studies 3 and 5 of the current programme incorporated 6-month maintenance tests, the results of which suggested that the benefits of relational training were indeed retained over this period. This adds further evidence in support of the utility of MET for relational framing and of the particular procedure used here for containment and hierarchical framing specifically.

The current research also adds to evidence showing that relatively sustained training (i.e., training provided over a relatively extended period of weeks or months) in relational framing may be able to boost intellectual performance. Previous research has shown that sustained training of relational framing of coordination, opposition and comparison in young children could positively impact intellectual performance (as measured on IQ and academic tests; see e.g., Cassidy et al., 2011, 2016; Hayes & Stewart, 2016). The current research is the first to provide similar evidence with respect to other framing repertoires, namely containment (Study 3) and hierarchy (Study 5); more specifically, it suggests (albeit tentatively) that training arbitrary containment and arbitrary hierarchy repertoires may result in improved performance on tests of categorisation, language and class inclusion. The numbers of participants employed in this context were small but the findings are suggestive at least. These findings add to evidence of the importance of targeting relational framing in general as a means of boosting general cognitive ability, and perhaps of targeting the frames trained here (containment, hierarchy) for improvement in particular skills such as categorisation or class inclusion. No firm conclusions may be drawn in the latter regard, however, until better controlled larger ‘n’ research is conducted. As such, it is recommended based on the current indicative findings that such follow-up work should be conducted.

Implications for Behaviour Analytic Research

As explained in the introduction, behaviour analytic researchers and practitioners have certainly recognised the importance of classification, and extensive basic and applied work has been conducted in this domain. For example, one recent stream of research has focused on assessing and training categorisation relevant skills using a derived equivalence paradigm (Kobari-Wright & Miguel, 2014; Miguel & Kobari-Wright, 2013; Miguel, Petursdottir & Carr, 2005; Miguel, Petursdottir, Carr & Michael, 2008; Petursdottir, Carr, Lechago & Almason, 2008). In these studies, which have involved both typically developing pre-schoolers as well as children with autism, participants have typically first been trained to tact and/or receptively identify a number of different though categorically related stimuli (e.g., wrenches, hammers and drills) both individually as well as by using a common category name (such as “tools”). Subsequently the children have been tested for emergent categorisation-type skills, such as being able to name members of the category (e.g., by responding correctly to the question “Tell me some tools”) or correctly naming the category when given names of members (e.g., responding “tools” when told “wrench, hammer and drill”) without additional training.

While this and other previous research on categorisation is important, it has focused almost exclusively on categorisation at a basic level while neglecting the hierarchical relational dimension of categories. However, if individuals fail to respond to the hierarchical relations between members and classes and between classes and other (e.g., higher or lower order) classes then they might simply be learning little more than to respond with more than one name or label for the same stimulus, or at best, to respond with different labels under contextual control. In other words, if they are not learning to respond in accordance with generalised hierarchical relations between stimuli then their understanding is limited. In more conventional language we might say that they fail to truly comprehend categories and categorisation as a whole.

The present work offers a functional analytic approach to hierarchical classification and its acquisition in terms of generalised patterns of containment and hierarchical relational framing which may be useful for researchers and practitioners within behaviour analysis. Study 1 provides some information as to the age at which relational framing repertoires relevant to hierarchical classification are acquired, supplying behaviour analysts with data regarding age-appropriate goals. This could also allow practitioners to identify deficits in responding on these relational repertoires based on a client’s performance on these relational tasks when compared to a same-aged peer. With regard to basic or applied work focusing on training categorisation, the protocols described that focused on training arbitrary containment and arbitrary hierarchy (Studies 2 – 5) might certainly help guide effective intervention to teach classification.

Furthermore, Studies 3 and 5 suggest that training these repertoires may provide some improvement on performance on standardised measures of language (i.e., PPVT-4) and categorisation (CCT), while also showing some impact on class inclusion responding. This kind of ready generalisation of the effects of training should be welcome in the applied arena. From the RFT point of view, as indeed already indicated, there is mounting evidence that training relational framing can boost linguistic and cognitive performance in general (see e.g., Cassidy et al., 2011, 2016; Hayes & Stewart, 2016) to which the current results add. However, most of this research including the present work has been with typically developing children and thus further work is needed with children with developmental delay for whom linguistic and cognitive deficits present a particular challenge.

Perhaps equally importantly, further research is needed to examine the effect of training particular frames on theoretically related areas of cognitive performance such as categorisation or classification, especially in populations in whom such repertoires may be deficient. A key purpose of this thesis has been to focus on understanding the establishment (either through exposure to naturalistic contingencies in the environment or through direct targeted training) of relational responding involved in classification. By gaining an understanding of the contingencies involved, what is learned when testing and training typically developing children can then be applied also with children and individuals with deficits in their relational repertoire. In the core studies of the present research programme, containment and hierarchical relations were extensively tested and trained in groups of typically developing children, providing insight into typical developmental acquisition of containment and hierarchy relations, their correlations with linguistic and cognitive development overall as well specifically with classification as a repertoire as well as information regarding the capacity to train children using relational framing techniques and some indication of the potential of such training. Due to limitations of time and resources this same assessment and training could not be provided to individuals with ASD but this is of key importance and thus this a critically important area for expansion in the future. However, at the least the success with typically developing children in this context suggests the potential of this work and provides some indicators regarding how it might be conducted.

In the case of "Piagetian" class inclusion meantime, the current programme (i.e., Studies 6-8) successfully used an RFT-based approach to training this repertoire in both individuals with developmental delay as well as typically developing children.

These findings arguably add to the considerable wealth of behaviour analytic research with developmentally disabled populations and also offer the first studies to assess and train class inclusion from a behaviour analytic perspective. These results suggest that the field has much to offer regarding the facilitation of developmental milestones with both typically developing and developmentally disabled populations. They also further underline the potential of RFT methods for facilitating complex repertoires. The concept of MET of a relational pattern under precise and consistent contextual control and utilising non-arbitrary guidance to facilitate the establishing of a more arbitrary pattern are all key hallmarks of an RFT approach and in this case allowed training, maintenance and generalisation of the derived relational pattern involved with very young typically developing children as well as with individuals with developmental delay including ASD. Such work will need extension and elaboration (for example, regarding exactly what other repertoires are needed to support participants for training in class inclusion) but the success of the current approach can at least help with respect to assessment and training of this repertoire.

Implications for Cognitive Developmental Psychology

As described in the introductory chapters, cognitive developmental research has predominantly focused on offering a descriptive analysis of hierarchical classification, by identifying the ages at which hierarchical classification repertoires of varying complexity 'emerge' in young children and adults.

In describing key research from this perspective in Chapter 1, I focused in particular on Blewitt (e.g., 1989, 1994). She suggests that children begin by distinguishing between objects based on physical appearance (based on size, shape, texture, etc.) and that they subsequently classify stimuli based on the formal properties of these stimuli; however, as they develop, they are increasingly able to relate to classes in a more abstract way, reasoning "about classes without referring to the surface characteristics or formal features of the stimuli in question". Translated into RFT language, the above suggests a movement from non-arbitrary to arbitrary relational responding.

The results of Study 1 are in agreement with this suggestion. Those findings indicated that non-arbitrary containment repertoires (i.e., those based on perceptual characteristics) were the first of these repertoires, acquired in early childhood, while the arguably more complex repertoires of arbitrary containment and arbitrary hierarchy are established later in childhood. Furthermore, the results of Study 1 indicated that these repertoires strengthened over time and that acquisition was strongly related to the age of the participant, providing further evidence to corroborate the perceptual-conceptual developmental trend proposed by cognitive developmental research such as that just cited. In addition to providing further evidence of the acquisition of classification repertoires in this manner (i.e., from perceptual to conceptual classification, or from non-arbitrary to arbitrary), Study 1 also provides a preliminary developmental profile, which provides some information as to the ages at which these repertoires emerge and become fully established. These data may supplement cognitive developmental theory and evidence regarding the age at which these repertoires are acquired.

Given the relationship between these relational repertoires (i.e., non-arbitrary containment, arbitrary containment, and arbitrary hierarchy) and performance on standardised measures of both language and cognition as evidenced by Study 1, it may be beneficial to train such responding. Given their focus on maturation as opposed to environmental control, cognitive developmental research has not tended to put emphasis on this. In contrast the present functional analytic approach would emphasise the importance of training the repertoires at issue and this was a key element of the work presented here. The studies outlined in Chapters 4 and 5 (i.e., Studies 2 – 5) demonstrated establishing and augmenting both arbitrary containment and arbitrary hierarchy relational repertoires in typically developing children. Such work should be of interest to cognitive developmental researchers as well as those in other traditions in showing how to explain and facilitate the emergence of these behaviours. Furthermore, Studies 3 and 5 suggest that training repertoires of arbitrary containment and arbitrary hierarchy may positively impact children's performance on standardised measures of language (i.e., PPVT-4) and categorisation (i.e., CCT), in addition to increasing the accurate performance of class inclusion tasks. This work supplements and extends the scope of the research offered by the cognitive developmental approach.

Study 6 within the present programme focused on assessing and training class inclusion responding in typically developing young children (ages 3 – 4). Class inclusion is a task that originally emerged from within the cognitive developmental paradigm (see e.g., Piaget, 1952). Cognitive developmental theory and previous research have suggested that this repertoire does not typically emerge at this age, but is established later in childhood (see Winer, 1980). The baseline assessment for all three participants in Study 6 indicated that, as cognitive developmental psychologists would predict, class inclusion responding was not yet established in these young children and performance was at chance level. However, findings showed that it was possible to teach this repertoire to this young age group. Findings showed not just acquisition but also generalisation and maintenance. This extends the findings of previous research from within cognitive developmental psychology since while some previous studies from within this paradigm had shown some success in training these repertoires in young children, none had successfully demonstrated both generalisation and maintenance.

Study 7 added further in this regard that should be of interest with respect to cognitive developmental research. This study appears to be the first to successfully assess and train class inclusion responding in individuals with a diagnosis of autism spectrum disorder (ASD). Despite each of the participants being older than those within Study 6 (8 years 1 month – 19 years 1 month), none of the participants demonstrated an established class inclusion repertoire. As ASD is a developmental disorder, it is not surprising that deficits in this repertoire, which has largely been regarded as a developmental milestone, were observed. The results of Study 7 indicated that class inclusion responding could be successfully trained in the population concerned. Study 8 also trained class inclusion responding in individuals with ASD (plus an additional diagnosis; e.g., intellectual disability, or learning disability) but advanced on Study 7 in several respects. For example, it employed a stronger research design (i.e., concurrent multiple baseline design) and data was also recorded regarding the stimuli assessed at baseline in order to more accurately gauge generalisation of class inclusion performance. Study 8 contributed to our current understanding of class inclusion performance among individuals with developmental disabilities and again indicated that class inclusion responding can be established in this population.

Collectively the studies in the present programme which assessed and trained hierarchical classification including class inclusion in a range of young children should be of interest to cognitive developmental researchers both because they have re-capitulated some of their key findings in terms of developmental sequence and because, by showing successful training of the repertoires at issue in very young and / or developmentally delayed populations that would not be predicted to be able to show them, they have gone beyond the maturational paradigm that characterises the cognitive developmental approach and provided data and procedures that should be of interest both to basic researchers investigating childhood development as well as applied practitioners concerned to find effective ways of teaching core skills to young children.

Clinical Implications of the Work

As previously outlined within the current chapter, Study 1 provides some information as to the age at which these repertoires naturally emerge, supplying practitioners with data regarding age-appropriate goals. Additionally, the protocols outlined may act as a preliminary guide to identify deficits in responding on these relational repertoires based on a client's performance on these relational tasks when compared to a same-aged peer. Studies 2 to 5 also provide practitioners with protocols, which may be suitable to remediate deficits with arbitrary containment and arbitrary hierarchical repertoires.

Studies 6 – 8 which focused on establishing class inclusion repertoires among both young typically developing children and individuals with ASD provides an effective protocol to facilitate this developmental milestone. The promising results generated in Studies 6 – 8 indicate that the field has much to offer regarding the facilitation of developmental milestones with both typically developing and developmentally disabled populations. Therefore, such a protocol could be packaged and introduced within educational curriculums in both mainstream and specialised schools to facilitate this repertoire.

Early intervention programmes have traditionally assessed the progress of their students with developmental disabilities by measuring their acquisition of content; for example the number of echoics, tacts, mands, or intraverbals in an individuals repertoire (e.g., *Verbal Behavior Milestones Assessment and Placement Program*; Sundberg, 2008). However, on the basis of findings presented in this thesis, particularly that of Study 1, the ability to readily demonstrate derived relational responding in accordance with hierarchy and containment appears to be critical to more advanced linguistic and cognitive performance among typically developing children. These findings further support the suggestion that a student's progress should be assessed systematically in regards to these critical language and cognitive learning processes. For example, if a practitioner has assessed a student and consequently knows the level of derived relational ability that he/she can reliably perform (e.g., mutually entailed containment relations) this information can direct a practitioner towards developing a highly efficient curriculum designed to facilitate the acquisition or expansion of a range of skills.

An educational curriculum formulated on this basis could also provide practitioners with the opportunity to capitalise on already established derived relational responding within the students' repertoire to facilitate or augment a number of socially and educationally meaningful skills (e.g., functional communication training, mathematical reasoning). For instance, Rehfeldt and Root (2005) capitalised on their participants' existing derived relational coordinate responding repertoire to expand the derived requesting skills of three adults with developmental disabilities. Within this procedure, participants were first trained to request for preferred items using pictures; participants were then taught conditional discriminations between pictures and their spoken names and between spoken names and their corresponding text. Following training, all three participants successfully demonstrated derived requesting repertoires by requesting preferred items using text. Similarly, Halvey and Rehfeldt (2005) indicate that training specific conditional discriminations can result in the expansion of derived vocal requesting skills. More recently, Dixon and colleagues (2017) taught three children with developmental disabilities three 4-member equivalence classes which were comprised of three stimuli (e.g., A, B, and C) and a category name (e.g., D) using a match-to-sample procedure. The results of this study indicated that all participants demonstrated derived categorical responding in addition to untrained intraverbal categorical responses. These studies successfully established requesting, categorical and intraverbal categorical repertoires via derived relational responding, thereby indicating that this may be an effective and efficient method of programming for the emergence of a number of new skills. Studies 2 – 5 within the current thesis illustrates an effective method for facilitating arbitrary containment and arbitrary hierarchy repertoires, while studies 3 and 5 indicate that teaching these derived relational responses may also impact language and cognitive performance.

Furthermore, such an educational curriculum would allow a behaviour analyst to determine the stage at which a student has demonstrated mastery of a particular relational skill and may act as an indicator of when to stop explicitly targeting specific skills and instead aim to identify new target skills for the student. For instance, if the results of such an assessment indicated that a student could reliably demonstrate combinatorially entailed coordinate relations, it would indicate that a more efficient education program would focus on teaching content such as nouns, verbs and so on via one mode (e.g., teaching as a listener discrimination) and then assess this content in other modes (e.g., assess as a tact). Such an educational curriculum could also potentially identify particular deficits in a students derived relational responding repertoires (e.g., deficits in mutually entailed containment responding), then an individualised curriculum program could potentially focus on establishing those skills through multiple exemplar training on the relevant pattern of responding – such as the procedures outlined in Studies 2 – 5 which focused on facilitating arbitrary containment and arbitrary hierarchy repertoires.

Issues and Directions for Future Research

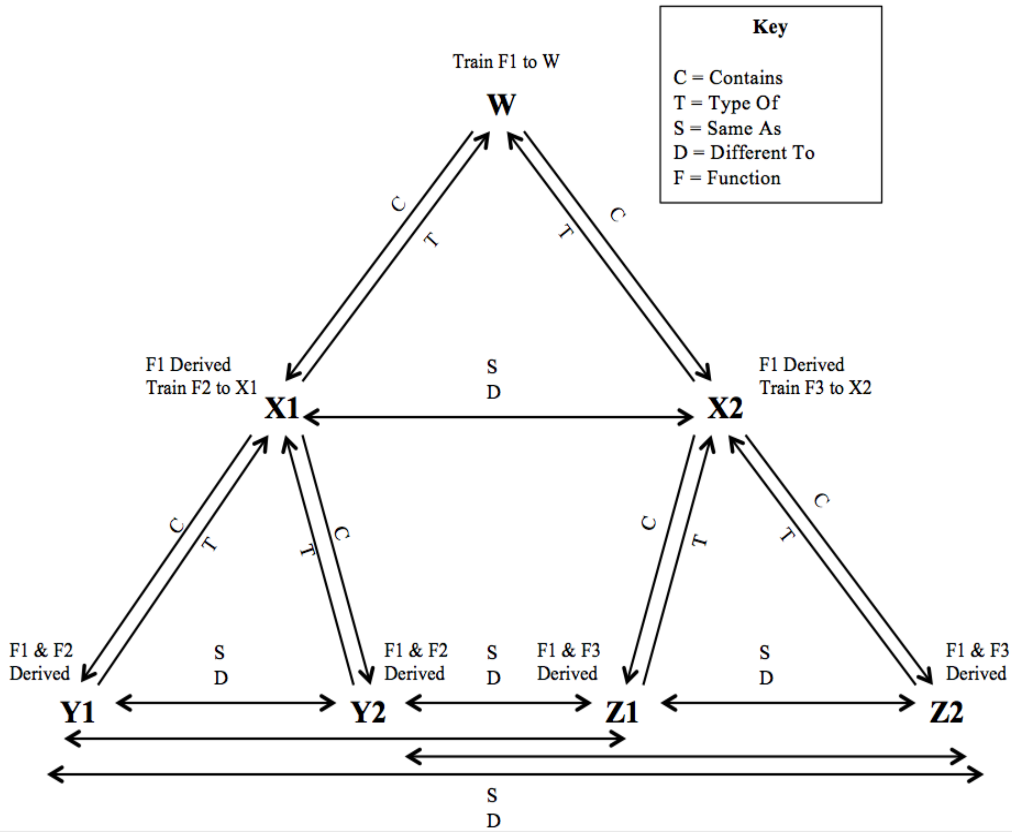
Although there were a number of strengths featured in the current work, a number of issues or limitations were also apparent which point to the importance of particular avenues for continuing empirical investigation.

Assessment of Hierarchical Classification

A core aim of the current research programme was to assess and train relational framing related to hierarchical classification including both containment and hierarchical relational framing. With respect to the latter in particular, it sought to assess properties of relational framing (i.e., mutual entailment, combinatorial entailment and transformation of stimulus function) that align with the key properties of hierarchical classification (i.e., transitive class containment, asymmetrical class containment and unilateral property induction) as expounded and tested in the mainstream literature. In doing so, the research assessed hierarchical classification using particular cues, including primarily “type of” and “contains”, which seemed particularly pertinent to this end and by testing and training children using this approach it has provided an arguably important advance. At the same time, however, hierarchical relational framing is a complex repertoire and recent RFT work including the current study is arguably only beginning to engage with this repertoire. As such, while research such as the present project is an important step forward, there remains considerable work yet to do in exploring and teaching this repertoire.

While the present research did indeed investigate key predicted patterns of derivation in accordance with hierarchical relations based on cues such as the above, it arguably did not examine the full gamut of derivations that might be part of a comprehensive repertoire of hierarchical relational framing and hierarchical classification. The presence of a more complex set of derivations is hinted at by the following quote from Hayes and colleagues (2001) who suggest that hierarchical relations share “the same basic relational pattern of a frame of comparison”, such that a hierarchical frame is “like a frame of comparison in the sense that it is diode-like but because it is not merely qualitative, the combinatorial relations tend to be more specific even without quantification” (p. 37). In fact, responding in accordance with hierarchical classification in a relatively comprehensive sense likely requires responding appropriately in accordance with a variety of derivable relations including for example coordination, distinction and comparison. This is because hierarchical classification in its fullest sense is a relational network that involves, for example, classes of stimuli that can be similar or different from each other depending on whether they are in the same or different classes (thus implying sameness and difference relations) and that also involves classes that can subsume other classes, thus implying classes that are bigger or smaller.

Figure 7.1. Example of hierarchical relational network.



Consider the hierarchical relational network represented in Figure 7.1, for example. In this network, the class W at the top of the hierarchy contains the subclasses X1 and X2, while each of these classes in turn is further subdivided such that X1 contains the stimuli Y1 and Y2 while X2 contains Z1 and Z2. When two stimuli (e.g., Y1 and Y2) are framed as being part of a particular class, they might further be derived as being the same as each other and derived as being different from other stimuli that have previously been framed as being in another class (e.g., Z1 and Z2). Deriving the latter, however, might depend on the nature and extent of the network involved. For example, if the class X2 in which Z1 and Z2 reside is framed as being similar or closely related to the class X1, then Y1 and Y2 may be derived as being similar to Z1 and Z2, albeit not as similar as they are to each other. Furthermore, if, say, Y1 acquires a particular function (e.g., it is evaluated as good) then the transformation of function of the other stimuli in the network might occur in accordance with the patterns just described as well as the hierarchical relations explored as part of the present research programme. For example, Y2 might also be evaluated as good, while Z1 and Z2 might be evaluated as being not as good, though with their level of goodness determined more precisely by how closely the Y and Z classes are to each other.

This kind of pattern of derivation and transformation of function might perhaps be predicted in someone with a sufficiently advanced hierarchical classification repertoire. As such, testing and training this more complex type of pattern should be seen as the ultimate target for researchers and practitioners. Further work on hierarchical classification in children might attempt to expand the hierarchical classification protocol used within the current programme of research so as to allow assessment and training of a more complex pattern of possible derived relations such as this. Successful work might be expected to improve the reliability and validity of the instrument used herein as an assessment and provide the basis of a much more thoroughgoing and generally beneficial training paradigm.

In addition to this, children's performance in accordance with hierarchical classification could also be more sharply delineated by comparing and contrasting this pattern of hierarchy with other forms of hierarchical responding. For example, as described in the introduction to the thesis, part-whole or analytic hierarchy (cues: part of, contains) is a form of hierarchical responding that would not seem to share the pattern of transformation of functions seen in hierarchical classification (cues: type/member of, contains) (Markman & Seibert, 1976; Slattery & Stewart, 2014). As previously described, in the case of hierarchical classification, functions trained in higher order stimuli will transform the functions of stimuli directly below them in the hierarchy but the reverse is not true (as described in Chapter 1, this is referred to in the mainstream literature as unilateral property induction). For example, if I am told that a ‘zag’ is a type of animal and that all animals breathe then I can derive that ‘zags’ breathe; however if I am told that a ‘zag’ has a particular characteristic then I cannot derive that all animals have that characteristic. While this pattern characterises hierarchical classification, however, it does not seem to characterise part/whole or analytic hierarchy. For example, if I am told that a ‘zog’ is part of a ‘car’ and that cars are vehicles that allow people to be transported, I cannot then derive that zogs are vehicles that allow people to be transported. If I am told that a ‘zog’ is primarily used for navigation (for example), I cannot conclude that cars are also primarily used for navigation. In other words, at the very least, part/whole (analytic) hierarchy does not seem to facilitate the same pattern of transformation of function as member/class (classification) hierarchy. Assessing and teaching this discrimination between the different patterns of hierarchy in terms of the expected pattern of transformation of function that might be expected could potentially be important in the assessment and training of each.

Expansion of the present assessment might involve also moving away from the ‘Yes’ and ‘No’ format upon which much of the current testing relied. Studies 1-5 of the current programme which assessed and trained non-arbitrary containment, arbitrary containment and arbitrary hierarchy relied solely on the participant providing “yes”, “no” or “don’t know” as a response. This type of format is a feature of the Relational Evaluation Procedure (REP), which has been used in a number of RFT research programmes to assess and train relational responding and it has been highly effective across these programmes (see e.g., Cassidy, Roche & Hayes, 2011) as, indeed, it has proven in the current programme also. However, particularly when dealing with an advanced and/or complex repertoire such as hierarchical relational framing, heavy reliance upon one particular format may not be ideal. The addition of other response formats might supplement the current format to facilitate a more nuanced assessment and/or training regime. For example, while ‘Yes’ and ‘No’ might be initial options on particular tasks, participants might also be asked to provide justification for their responses (i.e., explain why they gave the response they did), and more accurate or precise answers could be scored more highly than others. For example, responding that “A zimp is a type of wilk” because “I know that the class wilk contains yalts, and the class yalts contains zimps” might be scored more highly than simply answering ‘Yes’ or ‘No’ to a question regarding the relationship between ‘zimps’ and ‘wilks’.

Expanded testing and training of hierarchical relational framing underlying classification might incorporate other procedural variations also. One such might be presenting a child with an already established relational network involving familiar words and then introducing one or more novel (nonsense) stimuli into the network to test for appropriate patterns of derivation and transformation of function. For example, a child might be told that a ‘Zig’ is a type of dog and then asked questions about the relationship between a ‘Zig’ and other animals as well as about the characteristics of a ‘Zig’. If need be (i.e., to ensure their performance is not hampered through lack of familiarity with the relational network itself as well as to facilitate testing of performance in accordance with a relatively complex network), they might be presented with a diagram of the relational network involved before being asked questions. Indeed, perhaps more advanced participants might even be required to fill in or construct a hierarchical chart. For example, after being provided with information regarding the hierarchical network of a fictitious taxonomy (e.g., an alien kingdom) a participant might then be required to draw out a chart depicting the various levels of hierarchy within this abstract stimulus set.

Assessing the Impact of Training on Performance

One of the patterns seen in the current research programme was that training containment and hierarchical relational framing in Studies 3 and 5 respectively may have boosted intellectual performance including on tests of both general performance as well as classification specifically. However, as discussed in earlier chapters, the evidence provided was indicative only.

The primary design used in these studies was a single subject multiple baseline design across both participants and components. This design was chosen to show the efficacy of the training intervention, which it effectively did. A secondary aim of these studies, however, was to investigate the effect of the training on intellectual performance more generally. As such, in both studies a control group (n = 3) of participants who did not receive the training were assessed alongside the three training group participants on a number of standardised measures both before and after the training intervention.

The pattern of outcomes appeared relatively consistent across measures in showing a small boost for the training group on measures of both language and categorisation. Of course the number of participants involved was too small for conventional statistical testing and thus no reliable conclusion may be drawn.

However the data do provide some indication that training may be beneficial in this sense. In order to provide more substantive evidence a well-controlled large n group comparison design would be needed. Future research should endeavour to conduct a larger scale study of this kind in order to more adequately determine the possible influence of arbitrary containment and arbitrary hierarchical training on participants' performance on measures of language and categorisation.

Apart from this, a number of other avenues should also be pursued in future research on the benefits of training hierarchy. Such work might also employ longer and / or more intensive training. For example, in the present programme, the period of time over which the training was conducted in each of the intervention studies (i.e., Studies 2 – 8) was relatively brief (between 2 and 7 weeks) and training sessions, though typically delivered 4-5 times per week, were also quite short (between 15 and 20 minutes). Relational frame training studies using more time intensive training (i.e., sessions of 1 hour length) over longer time periods have previously been very successfully implemented for other (simpler) relations such as sameness, opposition and comparison. Given the success of such protocols (see, e.g., Cassidy et al., 2011, 2016; Hayes et al., 2016), it would be interesting and informative to conduct similar studies focused on hierarchy or other more complex relational repertoires.

Such future work might also employ better experimental controls. Certain controls were obviously present. For example, in all studies (i.e., 1 – 8), measures of inter-observer agreement (IOA) and procedural fidelity from an independent trained observer were in place. However, in other respects, aspects of data collection could be improved. For example, the primary investigator throughout was also the sole data collector. The administration of all tests and training protocols (i.e., in this case, the relational framing assessments in addition to the linguistic and cognitive measures) by the same researcher is a potential weakness as inadvertent prompting is often associated with table top procedures delivered by a single experimenter,

With respect to the between-group design, ideally, randomised controlled trials with blind testing should be implemented. In this respect also, the use of a single assessor across tests is less than desirable. The possibility of bias in favour of the experimental group is an obvious weakness within this context. The use of a single experimenter / tester was unfortunately unavoidable in the present context as the necessary resources to implement better control (e.g., employing multiple researchers) were not available. However, ideally, in future research a variety of different experimenters should be involved in the administration of the relational framing protocol and measures of cognitive functioning, such that a participant doesn't receive instruction or assessment from one experimenter across all phases of the experiment. In addition, with respect to assessment of participants across groups, ideally the assessor should be blind to condition. Such measures could offer a more reliable procedure for gauging the extent of a child's repertoire and/or the effect of the relational frame training.

Research examining the potential intellectual benefits of training hierarchy might also focus on more immediately academically relevant outcomes including school performance in general as well as within particular domains. Previous research, for example, has indicated that relational frame training can enhance children's academic attainment (e.g., Hayes & Stewart, 2016). Future research could examine the association between arbitrary containment and arbitrary hierarchical relational framing performances and mathematical attainment in both younger (e.g., using the Drumcondra Primary Mathematics Test – Revised; e.g., Hayes & Stewart, 2016) and older populations (e.g., Christmas and Summer Math Results or Junior Certificate Maths Results). Such work might in particular investigate the relationship between training in these frames and performance in math or science in which classification responding is perhaps particularly important and for which an established repertoire of framing relevant to hierarchical classification might prove particularly beneficial.

Future research on the effects of hierarchical relational training might extend on the present work in other respects also, perhaps by manipulating the level or comprehensiveness of the training provided. For example, a possible future study could compare a number of different groups before and after training. One group might serve as controls, receiving only the assessments before and after the intervention. A second group might receive training in arbitrary hierarchy alone, a third group (similar to the participants in the present research) might be trained in non-arbitrary and arbitrary containment as well as arbitrary hierarchy and a fourth group might be trained in the same relational frames as the third group but, in the case of arbitrary hierarchy targeting in addition the type of complexity discussed earlier on this chapter (and shown in Figure 7.1) including appropriate derivations and transformations of function in accordance with sameness, distinction and comparison relations. One might predict that all three of the relational frame training groups might demonstrate improvements in intellectual performance when compared with the control group. However, it might also be predicted that the relational framing groups might differ somewhat in this respect and possibly, more interestingly, that groups that received more comprehensive hierarchical training might show a correlated level of improvement in categorisation and classification performance specifically.

Finally, future research might also benefit from extending the range of participants trained and tested in relational framing of hierarchical classification. The current work targeting relational framing of containment and hierarchy (as opposed to work focused on class inclusion) focused on typically developing children within the range of 3-8 years exclusively. This was to some extent based on the outcomes of Study 1 which showed that this was the age range within which these frames seemed to be acquired over time and within which training might be most beneficial. However, this of course is merely a beginning in the investigation of this repertoire in relevant populations. Both younger and older children need to be assessed and trained in relevant repertoires. Younger children might be assessed and trained in the simplest repertoire seen here, namely, non-arbitrary containment. Older children might be tested and trained in particular on more advanced hierarchical relational framing such as discussed in the previous section. In addition to these groups, the current work should also be extended to children and adults with intellectual difficulties and in particular deficits in this relational domain. The present programme did train class inclusion in children with ASD but no work was done attempting to train containment or hierarchical relational framing for example. An extension of the work in this type of respect is potentially important especially with regard to the practical or applied utility of behaviour analysis and relational frame theory.

Summary and Conclusion

The core aim of the current doctoral research programme was to develop protocols for assessing and training relational framing of hierarchical classification in young children including both typically developing children and children with developmental delay. It pursued this aim over the course of a series of studies and has provided relevant data in this respect which have arguably contributed in several areas including RFT, on which it is directly based; behaviour analysis, the basic and applied science underlying RFT; and cognitive developmental psychology, within which most work in this domain had been conducted up until now. Summarising the current results again very briefly, Study 1 showed correlations between the relational framing repertoires at issue and measures of language cognition and categorisation; Studies 2 to 5 all used a multiple baseline design to show the efficacy of MET of these repertoires in young children and also provided some preliminary evidence of the potential impact of this training on language, categorisation and class inclusion performance; finally, Studies 6-8 used a multiple baseline design to show the efficacy of an RFT-based procedure for training ‘class inclusion’ in young typically developing children and individuals with a diagnosis of autism. This work has hopefully advanced existing behavioural science in terms of both its theoretical and practical understanding of hierarchical classification as a potentially important repertoire that is learned and that can be taught. However, as previously suggested, this constitutes only a preliminary investigation. The current programme of work has provided an important initial step and has also tried to indicate potential avenues for future research; the extent to which it has accomplished the latter in particular should be the benchmark of its success.

References

Adams, B. J., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. *The Psychological Record*, *43*, 553-566. Retrieved from: https://search.proquest.com/openview/625a87e8ab731447fbba7b8a9e7758bf/1?pq-origsite=gscholar&cbl=1817765

Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: training counteracts the representativeness heuristic. *Cognitive Development*, *6*. 195-217. DOI: https://doi.org/10.1016/0885-2014(91)90036-D

Ahr, P. R. & Youniss, J. (1970). Reasons for failure on the class-inclusion problem. *Child Development*, *41*. 131- 143. DOI: 10.2307/1127395

Alderson-Day, B., & McGonigle-Chalmers, M. (2011). Is it a bird? Is it a plane? Category use in problem-solving in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *41 (5)*, 555 – 565. DOI: https://doi.org/10.1007/s10803-010-1077-9

Arterberry, M. E., & Bornstein, M. H. (2012). Categorization of real and replica objects by 14- and 18-month-old infants. *Infant Behavior and Development*, *35*, 606 – 612. DOI: 10.1016/j.infbeh.2012.04.007

Arterberry, M. E., & Kellman, P. J. (2016). *Development of perception in infancy: The cradle of knowledge revistied*. Canada: Oxford University Press.

Astley, S. L., Peissig, J. J., & Wassermann, E. A. (2001). Superordinate categorization via learned stimulus equivalence: quantity of reinforcement, hedonic value and nature of the mediator. *Journal of Experimental*

Psychology: Animal Behavior Processes, *27 (3),* 252 – 268. DOI: http://psycnet.apa.org/doi/10.1037/0097-7403.27.3.252

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: evidence from exploratory play. *Child Development*, *64,* 711 – 728. DOI: 10.1111/j.1467-8624.1993.tb02938.x

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record*, *44,* 91-124. Retrieved from: https://search.proquest.com/openview/6e07d906308b5cc1f6d2dc69ad708197/1?pq-origsite=gscholar&cbl=1817765

Barnes, D., McCullagh, P., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *Analysis of Verbal Behavior*, *8,* 1 – 11. Retrieved from: https://link.springer.com/article/10.1007/BF03392844

Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Sinner's verbal behaviour: A possible synthesis. *The Behavior Analyst*, *23,* 69 – 84. DOI: https://doi.org/10.1007/BF03392000

Barnes-Holmes, Y., Barnes-Holmes, D., & McHugh, L. (2004). Teaching derived relational responding to young children. *Journal of Early and Intensive Behavior Intervention*, *1 (1),* 3 – 12. DOI: http://psycnet.apa.org/doi/10.1037/h0100275

Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P. M. (2004). Establishing relational responding in accordance with opposite as a generalized operant behaviour in young children. *International Journal of Psychology and*

Psychological Therapy, *4 (3),* 559 – 586. Retrieved from: http://www.redalyc.org/html/560/56040307/

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., & Luciano, C. (2004). A derived transfer of mood functions through equivalence relations. *The Psychological Record*, *54,* 95 – 113. Retrieved from: https://search.proquest.com/openview/1e557abb1a7dba0bf52d50828be75b86/1?pq-origsite=gscholar&cbl=41800

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy, 4,* 531-558. Retrieved from: http://www.redalyc.org/html/560/56040306/

Barnes-Holmes, D., Hayes, S.C., Dymond, S., & O'Hora, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. In S.C. Hayes, D., Barnes-Holmes & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 51-71). New York: Plenum.

Baron-Cohen, S., Joliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*, 813 – 822. DOI: 10.1111/j.1469-7610.1997.tb01599.x

Barrouillet, P. (1996). Transitive inferences from set-inclusion relations and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22 (6),* 1408 – 1422. DOI: http://psycnet.apa.org/doi/10.1037/0278-7393.22.6.1408

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for relational operants. *Journal of Applied Behavior Analysis, 40*, 45-71. DOI: 10.1901/jaba.2007.7-06

Berlin, B. (1972). Speculations on the growth of ethnobotanical nomenclature. *Journal of Language and Society*, *1,* 63-98. DOI: https://doi.org/10.1017/S0047404500006540

Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.

Best, J. (1995) *Cognitive psychology* (Eds.). Minneapolis: West.

Billow, R. M. (1975). A cognitive developmental study of metaphor comprehension. *Developmental Psychology, 11*. 415- 423. DOI: http://psycnet.apa.org/doi/10.1037/h0076668

Blewitt, P. (1983). Dog versus collie: vocabulary in speech to young children. *Developmental Psychology*, *19*. 602- 609. DOI: http://psycnet.apa.org/doi/10.1037/0012-1649.19.4.602

Blewitt, P. (1989). Categorical hierarchies: Levels of knowledge and skill. *Genetic Epistemologist*, *17,* 21-29.

Blewitt, P. (1994). Understanding categorical hierarchies: the earliest levels of skill. *Child Development*, *65,* 1279 – 1298. DOI: 10.1111/j.1467-8624.1994.tb00817.x

Boll, T. (1993). *Children's Category Test*. Psychological Corporation: San Antonio TX.

Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute, and group versus individual analyses. *Developmental Psychology*, *46*, 350 – 364. DOI: 10.1037/a0018411

Bornstein, M. H., Arterberry, M. E., & Mash, C. (2010). Infant object categorization transcends diverse object-context relations. *Infant Behavior and Development*, *33*, 7 – 15. DOI: 10.1016/j.infbeh.2009.10.003

Bornstein, M. H., & Mash, C. (2010). Experience-based and on-line categorization of objects in early infancy. *Child Development*, *81*, 884 – 897. DOI: 10.1111/j.1467-8624.2010.01440.x.

Borst, G., Poirel, N., Pineau, A., Cassotti, M., & Houdé, O. (2013). Inhibitory control efficiency in a Piaget-like class-inclusion task in school-age children and adults: a developmental negative priming study. *Developmental Psychology*, *49*, 1366 – 1374. DOI: http://psycnet.apa.org/doi/10.1037/a0029622

Brainerd, C. J. (1974). Training and transfer of transitivity, conservation, and class inclusion of length. Child Development, 45, 324 -334. DOI: 10.2307/1127952

Brino, A. L., Campos, R. S., Galvào, O. F., & McIlvane, W. J. (2014). Blank-comparison matching-to-sample reveals a false positive symmetry test in a capuchin monkey. *Psychology and Neuroscience*, *7 (2)*, 193 – 198. DOI: http://psycnet.apa.org/doi/10.3922/j.psns.2014.008

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York, NY: Wiley.

Byrnes, J. P., & Overton, W. F. (1986). Reasoning about certainty and uncertainty in concrete, causal and propositional contexts. *Developmental Psychology*, *22 (6)*, 793 – 799. DOI: http://psycnet.apa.org/doi/10.1037/0012-1649.22.6.793

California Department of Education (2008). *California preschool learning foundations (Vol. 1)*. Sacramento, CA: CDE Press.

Callanan, M. A. (1985). How parents label objects for young children: the role of input in the acquisition of category hierarchies. *Child Development, 56*. 508- 523. DOI: 10.2307/1129738

Canobi, K. H., Reeve, R. A., & Pattison, P. E. (2002). Young children's understanding of addition concepts. *Education Psychology: An International Journal of Experimental Educational Psychology*, *22 (5),* 513 – 532. DOI: http://dx.doi.org/10.1080/0144341022000023608

Carneiro, P., Albuquerque, P., & Fernandez, A. (2009). Opposite developmental trends for false recognition of basic and superordinate names. *Memory*, *17*, 411 – 427. DOI: 10.1080/09658210902758847

Carson, M. T., & Abrahamson, A. (1976). Some members are more equal than others: the effect of semantic typicality on class-inclusion performance. *Child Development*, *47 (4),* 1186 – 1190. DOI: 10.2307/1128461

Carpentier, F., Smeets, P. M., & Barnes-Holmes, D. (2003). Equivalence-equivalence as a model of analogy: Further analyses. *The Psychological Record*, *53,* 349-372. Retrieved from: https://search.proquest.com/openview/32cdd96b60c48842e5feeec663f38ed0/1?pq-origsite=gscholar&cbl=41800

Carson, M. T. & Abrahamson, A. (1976). Some members are more equal than others: the effect of semantic typicality on class-inclusion performance. *Child Development, 47*. 1186- 1190. DOI: 10.2307/1128461

Cassidy, S., Roche, B., Colbert, D., Stewart, I. & Grey, I. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences, 47,* 222-235. DOI: 10.1016/j.lindif.2016.03.001

Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise Intelligence Quotients: A pilot study. *The Psychological Record, 61,* 173–198. Retrieved from: https://search.proquest.com/openview/3c333000d49ea7dd7024a203b6fe352c/1?pq-origsite=gscholar&cbl=41800

Cassidy, S., Roche, B., & O'Hora, D. (2010). Relational Frame Theory and human intelligence. *European Journal of Behavior Analysis, 11,* 37–51. Retrieved from: http://www.tandfonline.com/doi/abs/10.1080/15021149.2010.11434333

Catania, A. C., Horne, P., & Lowe, C. F. (1989). Transfer of function across members of an equivalence class. *The Analysis of Verbal Behavior, 7 (1),* 99 – 110.

Christ, T. J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools, 44 (5),* 451-459. DOI:10.1002/pits.20237

Church, B. A., Krauss, M. S., Lopata, C., Toomey, J. A., Thomeer, M. L., Coutinho, M. V., Volker, M. A., & Mercado III, E. (2010). Atypical categorization in children with high-function autism spectrum disorder. *Psychonomic Bulletin and Review, 17 (6),* 862 – 868. DOI: 10.3758/PBR.17.6.862

Church, B. A., Rice, C. L., Dovgopoloy, A., Lopata, C. J., Thomeer, M. L., Nelson, A., & Mercado III, E. (2015). Learning, plasticity, and atypical generalization in children with autism. *Psychonomic Bulletin and Review, 22 (5)*, 1342 – 1348. DOI: 10.3758/s13423-014-0797-9

Coley, J. D., Medin, D. L., & Atran, S. (1997). Does rank have its privilege? Inductive inferences within folkbiological taxonomies. *Cognition*, *64*, 73 – 112. Retrieved from: https://www.researchgate.net/profile/John_Coley/publication/13885568_Does_rank_have_its_privilege_Inductive_inference_within_folkbiological_taxonomies/links/0c960518111b2a4493000000/Does-rank-have-its-privilege-Inductive-inference-within-folkbiological-taxonomies.pdf

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behaviour*, *8*, 240-247. DOI: https://doi.org/10.1016/S0022-5371(69)80069-1

Collins, A. M., & Quillian, M. R. (1972). How to make a language user. Cited in: Tulving, E., & Donaldson, W. (Eds.). *Organization of memory*. Oxford, England: Academic Press.

Community-University Partnership for the Study of Children, Youth, and Families (2011). *Review of the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4).* Edmonton, Alberta, Canada: Community University Partnership.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). Applied Behavior Analysis, 2nd ed. New York: Pearson Education.

Cowan, R., & Menton, M. (1996). Do they know what they are doing? Children's use of economical addition strategies and knowledge of commutativity.

Educational Psychology: An International Journal of Experimental Educational Psychology, *16 (4)*, 407 – 420. DOI: http://dx.doi.org/10.1080/0144341960160405

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis*, *25*, 461 – 475. DOI: 10.1901/jaba.1992.25-461

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, *48 (3)*, 497 – 507. DOI: https://doi.org/10.1016/j.neuron.2005.10.018

Deneault, J., & Ricard, M. (2005). The effect of hierarchical levels of categories on children's deductive inferences about inclusion. *International Journal of Psychology, 40*. 65- 79. DOI: 10.1080/00207590444000032

Deneault, J., & Ricard, M. (2006). The assessment of children's understanding of inclusion relations: transitivity, asymmetry, and quantification. *Journal of Cognition and Development*, *7 (4)*, 551 – 570. DOI: http://dx.doi.org/10.1207/s15327647jcd0704_6

Deng, W. S., & Sloutsky, V. M. (2015). The development of categorization: effects of classification and inference training on category representation. *Developmental Psychology*, *51 (3)*, 392 – 405. DOI: http://dx.doi.org/10.1037/a0038749

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. DOI: 10/1901/jeab.1986.46-243

Dickins, D. W., Singh, K. D., Roberts, N., Burns, P., Downes, J. J., Jimmieson, P.,

& Bentall, R. P. (2001). An fMRI study of stimulus equivalence. *NeuroReport, 12(2),* 405. Retrieved from: http://journals.lww.com/neuroreport/Abstract/2001/02120/An_fMRI_study_of_stimulus_equivalence.43.aspx

Diesendruck, G., & Shatz, M. (2001). Two year olds' recognition of hierarchies evidence from their interpretation of the semantic relation between object labels. *Cognitive Development, 16*. 577- 594. DOI: http://dx.doi.org/10.1016/S0885-2014(01)00047-8

Dixon, M. R., Belisle, J., Stanley, C. R., Speelman, R. C., Rowsey, K. E., Kime, D., & Daar, J. (2017). Establishing derived categorical responding in children with disabilities using the PEAK-E curriculum. *Journal of Applied Behavior Analysis, 50 (1),* 134 – 145. DOI: 10.1002/jaba.355

Dougher, M. J., Auguston, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The selective impact of Skinner's Verbal Behavior on empirical research: A reply to Schlinger (2008). *The Psychological Record, 60,* 353 – 360.

Dugdale, N., & Lowe, C. F. (1990). Naming and stimulus equivalence. Cited in: Blackman, D. E., & Lejeune, H. (Eds.). *Behavior analysis in theory and practice: Contributions and controversies*. Brighton, UK: Lawrence Erlbaum Associates.

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73,* 5 – 22. DOI: 10/1901/jeab/2000/73-5

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, 4th edition (PPVT-4)*. Pearson Education: Minneapolis: MN.

Dymond, S., & Roche, B. (2013). *Advances in relational frame theory: Research and application*. Oakland, CA: New Harbinger.

Edwards, D. J., Perlman, A., & Reed, P. (2012). Unsupervised categorization in a sample of children with autism spectrum disorders. *Research in Developmental Disabilities*, *33 (4)*, 1264 – 1269. DOI: https://doi.org/10.1016/j.ridd.2012.02.021

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. DOI: 10/1901/jeab.1992.58-123

Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903 – 917. DOI: 10.2307/1131427

Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, *58*, 418 – 431. DOI: http://dx.doi.org/10.1006/jecp.1994.1043

Eysenck, M. W., & Keane, M. T. (2000). *Cognitive Psychology: A student's handbook* (Eds.). USA: Psychology Press.

Fiebelkorn, I. C., Foxxe, J. J., McCourt, M. E., Dumas, K. N., Molholm, S. (2013). Atypical category processing and hemispheric asymmetries in high-functioning children with autism: Revealed through high-density

EEG mapping. *Cortex*, *49 (5)*, 1259 – 1267. DOI: https://doi.org/10.1016/j.cortex.2012.04.007

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. Retrieved from: https://search.proquest.com/openview/787457abeaa769632788e07b8026363c/1?pq-origsite=gscholar&cbl=1817765

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 (3)*, 345 – 358. DOI: 10.1901/jeab.1990.53-345

Fields, L., & Garruto, M. (2009). Optimising linked perceptual class formation and transfer of function. *Journal of the Experimental Analysis of Behavior*, *91 (2)*, 225- 251. DOI: 10.1901/jeab.2009.91-225

Fields, L., & Moss, P. (2008). Formation of partially and fully elaborated generalized equivalence classes. *Journal of the Experimental Analysis of Behavior*, *90*, 135- 168. DOI: https://dx.doi.org/10.1901%2Fjeab.2008.90-135

Fields, L., & Reeve, K. F. (2000). Synthesizing equivalence classes and natural categories from perceptual and relational classes. In J. C. Leslie & D. Blackman (Eds.), *Experimental and Applied Analysis of Human Behaviour* (pp. 59-83). Reno, Nevada: Context Press.

Fields, L., Reeve, K., Adams, B. J., Brown, J. L., & Verhave, T (1997). Predicting the extension of equivalence classes from the primary generalization gradients: the merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior*, *68*, 67-92. DOI: https://dx.doi.org/10.1901%2Fjeab.1997.68-67

Fields, L., & Watanabe-Rose, M. (2008). Nodal structure and the partitioning of equivalence classes. *Journal of the Experimental Analysis of Behavior*, *89 (3)*, 359 – 381. DOI: https://dx.doi.org/10.1901%2Fjeab.2008-89-359

Frank, A. J., & Wasserman, E. A. (2005). Associative symmetry in the pigeon after successive matching-to-sample training. *Journal of the Experimental Analysis of Behavior*, *84*, 147 – 165. FOI: 10.1901/jeab.2005.115-04

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cotext. *Nature*, *44*3, 85 – 88. DOI: 10.1038/nature05078

Furrer, S. D., & Younger, B. A. (2008). The impact of specific prior experiences on infants' extension of animal properties. *Developmental Science*, *11*, 712 – 721. doi: 10.1111/j.1467-7687.2008.00721.x

Galizio, M., Stewart, K. L., & Pilgrim, C. (2001). Clustering in artificial categories: An equivalence analysis. *Psychonomic Bulletin & Review*, *8*, 609-614. DOI: 10.3758/BF03196197

Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Typicality effects in contingency shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behavior*, *82*, 253-273. DOI: https://dx.doi.org/10.1901%2Fjeab.2004.82-253

Gash, H. (1982). Associated structures: class inclusion and role-taking. *The Journal of Genetic Psychology: Research and Theory on Human*

Development, *141 (2)*, 155 – 166. DOI: http://dx.doi.org/10.1080/00221325.1982.10533469

Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2014). Multiple baseline and multiple probe designs. In: Gast, D. L., & Ledford, J. R. (Eds.). *Single case research methodology: Applications in special education and behavioral sciences. 2nd ed*. London: Routledge. pp. 251 – 296.

Gastgeb, H. Z., Dundas, E. M., Minshew, N. J., & Strauss, M. S. (2012). Category formation in autism: Can individuals with autism form categories and prototypes of dot patterns? *Journal of Autism and Developmental Disorders, 42 (8)*, 1694 – 1704. DOI: https://doi.org/10.1007/s10803-011-1411-x

Gastgeb, H. Z., Rump, K. M., Best, C. A., Minshew, N. J., & Strauss, M. S. (2009). Prototype formation in autism: Can individuals with autism abstract facial prototypes? *Autism Research, 2 (5)*, 279 – 284. DOI: 10.1002/aur.93

Gastgeb, H. Z., Wilkinson, D. A., Minshew, N. J., & Strauss, M. S. (2011). Can individuals with autism abstract prototypes of natural faces? *Journal of Autism and Developmental Disorders, 41 (12)*, 1609 – 1618. DOI: https://doi.org/10.1007/s10803-011-1190-4

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65 – 95. DOI: https://doi.org/10.1016/0010-0285(88)90025-4

Gelman, S. A. (2003). *The essential child: origins of essentialism in everyday thought*. Oxford: University Press.

Gelman, S. A., & Meyer, M. (2011). Child categorization. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 95 – 105. DOI: 10.1002/wcs.96.

Gil, E., Luciano, C., Ruiz, F. J., & Valdiva-Salas, S. (2012). A preliminary demonstration of transformation of functions through hierarchical relations. *International Journal of Psychology and Psychological Therapy*, *12*. 1- 19. Retrieved from: http://www.redalyc.org/html/560/56023335002/

Gil, E., Luciano, C., Ruiz, F. J., & Valdiva-Salas, S. (2014). A further experimental step in the analysis of hierarchical responding. *International Journal of Psychology and Psychological Therapy*, *14 (2)*, 137 – 153. Retrieved from: http://www.redalyc.org/html/560/56031293001/

Gil-Luciano, B., Ruiz, F. J., Valdivia-Salas, S., & Suárez-Falcón, J. C. (2017). Promoting psychological flexibility on tolerance tasks: Framing behavior through deictic/hierarchical relations and specifying augmental functions. *The Psychological Record*, *67 (1)*, 1 – 9. DOI: http://dx.doi.org/10.1007%2Fs40732-016-0200-5

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of “sameness” and “difference” in an insect. *Nature, 410*, 930 – 933. Retrieved from: https://search.proquest.com/openview/63a3a1d29eac884ec1293a0f046ded eb/1?pq-origsite=gscholar&cbl=40569

Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General, 117 (3)*, 227 – 247. DOI: http://psycnet.apa.org/doi/10.1037/0096-3445.117.3.227

Gore, N. J., Barnes-Holmes, Y., & Murphy, G. (2010). The relationship between intellectual functioning and relational perspective-taking. *International Journal of Psychology and Psychological Therapy*, *10,* 1 – 17. Retrieved from: http://www.redalyc.org/html/560/56017066001/

Gorham, M., Barnes-Holmes, Y., Barnes-Holmes, D., & Berens, N. (2009). Derived comparative and transitive relations in young children with and without autism. *The Psychological Record, 59,* 221-246. Retrieved from: https://link.springer.com/article/10.1007/BF03395660

Goswami, U. & Pauen, S. (2005). The effects of a “family” analogy on class inclusion reasoning by young children. *Swiss Journal of Psychology*, *64*. 115- 124. DOI 10.1024/1421-0185.64.2.115

Greco, C., Hayne, H., & Rovee-Collier, C. (1990). Roles of function, reminding, and variability in categorization by 3-month-old infants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 617 – 633. DOI: http://psycnet.apa.org/doi/10.1037/0278-7393.16.4.617

Greene, T. R. (1991). Text manipulations influence children’s understanding of class inclusion hierarchies. *Journal of Experimental Child Psychology*, *52*. 354- 374. DOI: https://doi.org/10.1016/0022-0965(91)90069-5

Greene, T. R. (1994). What kindergartners know about class containment hierarchies. *Journal of Experimental Child Psychology*, *57,* 72-88. DOI: http://dx.doi.org/10.1006/jecp.1994.1004

Griffee, K., & Dougher, M. J. (2002). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior*, *78 (3)*, 433 – 447. DOI: https://dx.doi.org/10.1901%2Fjeab.2002.78-433

Hajibayova, L. (2013). Basic-level categories: A review. *Journal of Information Science*, *39*, 676 – 687. DOI: 10.1177/0165551513481443

Halvey, C., & Rehfeldt, R. A. (2005). Expanding vocal requesting repertoires via relational responding in adults with severe developmental disabilities. *The Analysis of Verbal Behavior*, *21 (1)*, 13 – 25. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2774094/pdf/anvb-21-01-13.pdf

Harmon, K., Strong, R., & Pasnak, R. (1982). Relational responses in tests of transposition with rhesus monkeys. *Learning and Motivation*, *13*, 495 – 504. DOI: https://doi.org/10.1016/0023-9690(82)90006-6

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hayes, S. C., Fox, E., Gifford, E., Wilson, K., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behaviour. In: Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). *Relational Frame Theory: A post-Skinnerian account of human language and cognition* (p. 21-50). New York, NY: Plenum Press.

Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *British Journal of Educational Psychology*, *86*, 397 – 411. DOI: 10.1111/bjep.12114

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*, 207 – 227. DOI: 10/1901/jeab.2000.74-207

Hendricks, C., & Pasnak, R. (1999). Effects of instruction in sequencing and class inclusion for first graders. *Genetic, Social and General Psychology Monographs*, *125*, 297-312. Retrieved from: https://search.proquest.com/openview/c10af8c0ee648e70aae09f9a05bd1553/1?pq-origsite=gscholar&cbl=36144

Hendricks, C., Pasnak, R., Willson-Quayle, A., Trueblood, L., Malabonga, V., & Ciancio, D. (1999). Effects of instruction in sequencing and class inclusion for first graders. *Genetic, Social, and General Psychology Monographs*, *125 (3)*, 297 – 312. Retrieved from: https://search.proquest.com/openview/c10af8c0ee648e70aae09f9a05bd1553/1?pq-origsite=gscholar&cbl=36144

Herold, R., Tényi, T., Lénárd, K., & Trixler, M. (2002). Theory of mind deficit in people with schizophrenia during medicine. *Psychological Medicine, 32*, 1125 – 1129. DOI: http://dx.doi.org/10.1017/S0033291702005433

Hodkin, B. (1987). Performance model analysis in class inclusion: An illustration with two language conditions. *Developmental Psychology*, *23*. 683–689. DOI: http://psycnet.apa.org/doi/10.1037/0012-1649.23.5.683

Hooper, F. H., Sipple, T. S., Goldman, J. A., & Swinton, S. S. (1979). A cross-sectional investigation of children's classificatory abilities. *Genetic Psychology Monographs*, *99*. 41- 89. Retrieved from: http://psycnet.apa.org/record/1979-25639-001

Hooper, F. H., Wanska, S. K., Peterson, G. W., & De Frain, J. (1979). The efficacy of small group instructional activities for training Piagetian concrete operations concepts. *The Journal of genetic Psychology: Research and theory on Human Development*, *134 (2)*, 281 – 297. DOI:

http://dx.doi.org/10/1080/00221325.1979.10534062

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65(1)*, 185-241. DOI:10.1901/jeab.1996.65-185

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Council for Exceptional Children*, *71 (2)*, 165 – 179.

Horobin, K., & Acredolo, C. (1989). The impact of probability judgments on reasoning about multiple possibilities. *Child Development*, *60*, 183–200. DOI: 10.2307/1131084

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*, 708 – 719. DOI: 10.2307/1129456

Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory: the basic account. In: Zettle, R. D., Hayes, S. C., Barnes-Holmes, D., & Biglan, A. (2016). *The Wiley handbook of contextual behavioral science*. Chichester: John Wiley & Sons Ltd.

Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory: implications for the study of human language and cognition. In: Zettle, R. D., Hayes, S. C., Barnes-Holmes, D., & Biglan, A. (2016). *The Wiley handbook of contextual behavioral science*. Chichester: John Wiley & Sons Ltd.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. New York: Norton.

Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development*, *75 (6)*, 1871 – 1885. DOI: 10.1111/j.1467-8624.2004.00822.x

Johnson, C. R., & Rakison, D. H. (2006). Early categorization of animate/inanimate concepts in young children with autism. *Journal of Developmental and Physical Disabilities*, *18 (2)*, 73 – 89. DOI: https://doi.org/10.1007/s10882-006-9007-7

Johnson, K. E., Scott, P., & Mervis, C. B. (1997). Development of children's understanding of basic subordinate inclusion relations. *Developmental Psychology*, *33,* 745 – 763. DOI: 10.1037/0012-1649.33.5.745

Judd, S. A. & Mervis, C. N. (1979). Learning to solve class-inclusion problems: the roles of quantification and recognition of contradiction. *Child Development*, *50*. 163- 169. DOI: 10.1111/1467-8624.ep12432800

Kalish, C. W., & Gelman, S. A. (1992). On wooden pillows: multiple classification and children's category-based inductions. *Child Development*, *63*, 1536 – 1557. DOI: 10.1111/j.1467-8624.1992.tb01713.x

Keller, F., & Schoenfeld, W. (1950). *Principles of psychology*. New York: Appleton-Century-Crofts.

Kilroe, H., Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2014). Using the T-IRAP interactive computer programme and applied behaviour analysis to teach relational responding in children with autism. *Behavioral Development Bulletin, 19,* 60 - 81. DOI: http://dx.doi.org/10.1037/h0100578

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology, 13*, 111 – 124. Retrieved from:

https://www.cambridge.org/core/journals/development-and-psychopathology/article/prototype-formation-in-autism/74098CA2709B96507280E30696F236BF

Kobari-Wright, V. V., & Miguel, C. F. (2014). The effects of listener training on the emergence of categorization and speaker behavior in children with autism. *Journal of Applied Behavior Analysis, 47(2),* 431-436. DOI:10.1002/jaba.115

Kofsky, E. (1966). A scalogram study of classificatory development. *Child Development*, *37*. 191- 204. DOI: 10.2307/1126439

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lane, S., D., Clow, J. K., Innis, A., & Critchfield, T. S. (1998). Generalization of cross modal stimulus equivalence classes: Operant processes as components in human category formation. *Journal of the Experimental Analysis of Behavior*, *70*, 267 - 279. DOI: https://dx.doi.org/10.1901%2Fjeab.1998.70-267

Langford, P. E. (1981). A longitudinal study of children's understanding of logical laws in arithmetic and Boolean algebra. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *1 (2)*, 119 – 139. DOI: http://dx.doi.org/10.1080/0144341810010202

Lazar, R. M., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in children. *Journal of the Experimental Analysis of Behavior*, *41*, 251 – 266. DOI: 10/1901/jeab.1984.41-251

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3 – 28. DOI: 10.1037//0096-3445.130.1.3

Lionello-DeNolf, K. M. (2009). The search for symmetry: 25 years in review. *Learning and Behavior*, *37,* 188 – 203. DOI: https://doi.org/10.3758/LB.37.2.188

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology, 56,* 201-239. DOI: 10.1006/jecp.1993.1032.

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: universal and cultural features of folk biological taxonomies and inductions. *Cognitive Psychology, 32,* 251 – 295. DOI: https://doi.org/10.1006/cogp.1997.0651

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological Review*, *111 (2),* 309 – 332. DOI: 10.1037/0033-295X.111.2.309

Lovell, K., Mitchell, B., & Everett, I .R. (1962). An experimental study of the growth of some logical structures. *British Journal of Psychology*, *53*. 175-188. DOI: 10.1111/j.2044-8295.1962.tb00824.x

Luciano, C., Gómez-Becerra, I., & Rodríguez-Valverde, M. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior, 87,* 349-365. DOI: 10.1901/jeab.2007.08-06

MacNeill, A. (1996). Book and Test Reviews. *Archives of Clinical Neuropsychology*, *11,* 171 – 173.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 3- 36. DOI: 10.1207/S15327647JCD0101N_2

Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, *23*. 263–298. DOI: http://dx.doi.org/10.1016/0010-0285(91)90011-C

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*. 291–318. DOI: http://dx.doi.org/10.1016/S0885-2014(93)80003-C

Margolis, E., & Laurence, S. (2000). *Concepts: Core Readings*. Cambridge, MA: The MIT Press.

Markman, E. M. (1989). *Categorization and naming in children: Problems in induction*. Cambridge: MIT Press.

Markram, K., & Markram, H. (2010). The intense world theory – a unifying theory of the neurobiology of autism. *Frontiers in Human Neuroscience*, *4*, 224. DOI: https://dx.doi.org/10.3389%2Ffnhum.2010.00224

Markman, E. M., & Seibert, J. (1976). Classes and collections: Internal organization and resulting holistic properties. *Cognitive Psychology*, *8*. 561–577. DOI: http://dx.doi.org/10.1016/0010-0285(76)90019-0

Matos, M. A., & d'Oliveira, M. M. H. (1992). Equivalence relations and reading. Cited in Hayes, S. C., & Hayes, L. J. (Eds.). *Understanding verbal relations*. Reno, NV: Context Press.

McCabe, A. E. & Siegel, L. S. (1987). The stability of training effects in young children's class inclusion reasoning. *Merrill-Palmer Quarterly*, *33*. 187-194. Retrieved from http://www.jstor.org/discover/10.2307/23086328?uid=3738232&uid=2&uid=4&sid=21102457621977

McGarrigle, J., Grieve, R., & Hughes, M. (1978). Interpreting inclusion: A contribution to the study of the child's cognitive and linguistic development. *Journal of Experimental Child Psychology, 26*. 528–550. DOI: http://dx.doi.org/10.1016/0022-0965(78)90131-5

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *The Psychological Record, 54(1),* 115-144. Retrieved from: https://search.proquest.com/openview/d131a4af0fac739455a4a5867179b315/1?pq-origsite=gscholar&cbl=41800

McHugh, L., Barnes-Holmes, Y., Barnes-Holmes, D., & Stewart, I. (2006). Understanding false belief as generalized operant behavior. *Psychological Record, 56*(3), 341-364. Retrieved from: https://search.proquest.com/openview/180a0cc43f06dad47bd16041eb2c915a/1?pq-origsite=gscholar&cbl=41800

McHugh, L., Barnes-Holmes, Y, Barnes-Holmes, D., Whelan, R., & Stewart, I. (2007). Knowing me knowing you: deictic complexity in false belief understanding. *Psychological Record, 57,* 533 – 542. Retrieved from: https://search.proquest.com/openview/2da4e94b6e69464dc4477bde009996c2/1?pq-origsite=gscholar&cbl=41800

McHugh, L., & Reed, P. (2008). Using Relational Frame Theory to build grammar in children with Autistic Spectrum Conditions. *The Journal of Speech-Language Pathology and Applied Behavior Analysis, 2.4-3.1,* 60-77 DOI: http://psycnet.apa.org/doi/10.1037/h0100233

McLay, L., Church, J., & Sutherland, D. (2014). Variables affecting the emergence of untaught equivalence relations in children with and without autism. *Developmental Neurorehabilitation*. *19 (2)*, 75 – 87. DOI: 10.3109/17518423.2014.899649

Meadows, S. (1977). An experimental investigation of Piaget's analysis of class inclusion. *British Journal of Psychology, 68*. 229- 235. DOI: 10.1111/j.2044-8295.1977.tb01580.x

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology*, *32 (1)*, 49 – 96. Retrieved from: https://doi.org/10.1006/cogp.1997.0645

Mercado, E., Church, B. A., Coutinho, M. V. C., Dovgopoly, A., Lopata, C. J., Toomey, J. A., & Thomeer, M. L. (2015). Heterogeneity in perceptual category learning by high functioning children with autism spectrum disorder. *Frontiers in Integrative Neuroscience, 9*, 42. DOI: 10.3389/fnint.2015.00042

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate level categories. *Child Development*, *53*. 258–266. DOI: 10.2307/1129660

Mervis, C. B., Johnson, K. E., & Mervis, C. A. (1994). Acquisition of subordinate categories by 3-year-olds: The role of attribute salience, linguistic input, and child characteristics. *Cognitive Development*, *9*. 211–234. DOI: http://dx.doi.org/10.1016/0885-2014(94)90004-3

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32,* 89-115. DOI:

https://doi.org/10.1146/annurev.ps.32.020181.000513

Miguel, C. F., & Kobari-Wright, V. V. (2013). The effects of tact training on the emergence of categorization and listener behavior in children with autism. *Journal of Applied Behavior Analysis, 46 (3)*, 669-673. DOI:10.1002/jaba.62

Miguel, C. F., Petursdottir, A. I., & Carr, J. E. (2005). The effects of multiple-tact and receptive-discrimination training on the acquisition of intraverbal behavior. *The Analysis of Verbal Behavior*, *27-41*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2774100/.

Miguel, C. F., Petursdottir, A. I., Carr, J. E., & Michael, J. (2008). The role of naming in stimulus categorization by preschool children. *Journal of the Experimental Analysis of Behavior*, *89 (3)*, 383-405. DOI:10.1901/ jeab.2008-89-383

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational ability and language performance in children with autism spectrum disorders and typically developing children: a further test of the TARPA protocol. *Psychological Record*, *64*, 233 – 251. DOI: 10.1007/s40732-014-0032-0

Moran, L., Walsh, L., Stewart, I., McElwee, J., & Ming, S. (2015). Correlating derived relational responding with linguistic and cognitive ability in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *19*, 32 – 43. DOI: 10.1016/j.rasd.2014.12.015

Mulhern, T. & Stewart, I. (2014, September). *Training class inclusion responding in young children*. Paper presented at the annual conference of the European Association for Behaviour Analysis, Stockholm, Sweden.

Mulhern, T., Stewart, I., & McElwee, J. (2017). Investigating relational frame of categorization in young children. *The Psychological Record*. DOI: 10.1007/s40732-017-0255-y

Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: The MIT Press.

Murphy, G. L. (2016). Is there an exemplar theory of concepts? *Psychonomic Bulletin & Review*, *23 (4)*, 1035 – 1042.

Murphy, C., & Barnes-Holmes, D. (2009). Derived more-less relational mands in children diagnosed with autism. *Journal of Applied Behavior Analysis*, *42 (2)*, 253 – 268. DOI: 10.1901/jaba.2009.42-253

Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2005). Derived manding in children with autism: synthesizing Skinner's verbal behaviour with relational frame theory. *Journal of Applied Behavior Analysis, 38 (4),* 445 – 462. DOI: 10.1901/jaba.2005.97-04

Naigles, L. R., Kelley, E., Troyb, E., & Fein, D. (2013). Residual difficulties with categorical induction in children with a history of autism. *Journal of Autism and Developmental Disorders*, *43 (9)*, 2048 – 2061. DOI: https://doi.org/10.1007/s10803-012-1754-y

Needham, A., Cantlon, J. F., & Ormsbee Holley, S. M. (2006). Infants' use of category knowledge and object attributes when segregating objects at 8.5 months of age. *Cognitive Psychology*, *53*, 345 – 360. DOI: 10.1016/j.cogpsych.2006.05.003

Newsome, K. B., Berens, K. N., Ghezzi, P. M., Aninao, T., & Newsome, W. D. (2014). Training relational language to improve reading comprehension. *European Journal of Behavior Analysis*, *15(2)*, 47-80. DOI: http://www.tandfonline.com/doi/abs/10.1080/15021149.2014.11434512

Ogawa, A., Yamazaki, Y., Ueno, K., Cheng, K., & Iriki, A. (2009). Neural correlation of species-typical illogical cognitive bias in human inference. *Journal of Cognitive Neuroscience, 22 (9)*, 2120 – 2130. DOI: http://dx.doi.org/10.1162/jocn.2009.21330

O'Hora, D., Pelaez, M., Barnes-Holmes, D., Rae, G., Robinson, K. & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, *58*, 569-584. Retrieved from: https://search.proquest.com/openview/837507939e16fec78ccf247e59b67ae5/1?pq-origsite=gscholar&cbl=41800

O'Toole, C., Barnes-Holmes, Murphy, C., O'Connor, J., & Barnes-Holmes, Y. (2009). Relational flexibility and human intelligence: Extending the remit of Skinner's verbal behaviour. *International Journal of Psychology and Psychological Therapy*, *9 (1)*, 1- 17. Retrieved from: http://www.redalyc.org/html/560/56012876001/

Palmer, D. C. (2002). Psychological Essentialism: A review of E. Margolis and S. Laurence (Eds.), Concepts: Core Readings. *Journal of the Experimental Analysis of Behavior*, *78*, 597-607. DOI: 10.1901/jeab.2002.78-597

Parra, I., & Ruiz, F. J. (2016). The effect on Intelligence Quotient of Training Fluency in Relational Frames of Coordination. *International Journal of Psychology and Psychological Therapy 16(1)* 1-12. Retrieved from: http://www.redalyc.org/html/560/56046596001/

Pasnak, R., Cooke, W. D., & Hendricks, C. (2006). Enhancing academic performance by strengthening class-inclusion reasoning. *The Journal of*

Psychology: Interdisciplinary and Applied, *140 (6)*, 603 – 613. DOI: http://dx.doi.org/10.3200/JRLP.140.6.603-613

Paul, R. (2008). Interventions to improve communication. *Child and Adolescent Psychiatric Clinics of North America*, *17 (4)*, 835 – x. DOI: 10.1016/jchc.2008.06.011

Pellegrini, A. D. (1983). Facilitating class-inclusion skills in school-age children. *The Journal of Genetic Psychology*, *143*. 29- 37. DOI:10.1080/00221325.1983.10533530

Perreault, A., Habak, C., Lepore, F., Mottron, L., & Bertone, A. (2015). Behavioral evidence for a functional link between low- and mid-level visual perception in the autism spectrum. *Neuropsychologia*, *77*, 380 – 386. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2015.09.022

Petursdottir, A. I., Carr, J. E., Lechago, S. A., & Almason, S. M. (2008). An evaluation of intraverbal training and listener training for teaching categorization skills. *Journal of Applied Behavior Analysis*, *41(1)*, 53-68. DOI:10.1901/jaba.2008.41-53

Piaget, J. (1952). *The child's conception of number*. New York: Norton.

Piérault-Le Bonniec, G. (1980). *The development of modal reasoning: Genesis of necessity and possibility notions*. New York: Academic.

Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding of inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception and guessing. *Developmental Psychology*, *36*, 169–179. DOI: 10.1037//0012-1649.36.2.169

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39 (5)*, 765 – 775. Retrieved from: https://www.cambridge.org/core/journals/journal-of-child-psychology-and-psychiatry-and-allied-disciplines/article/enhanced-discrimination-of-novel-highly-similar-stimuli-by-adults-with-autism-during-a-perceptual-learning-task/E0BBECFDC462547D93DF373FE557ED97

Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: The contribution of parental labelling and infants' categorization abilities. *Journal of Child Language*, *22*. 325–343. DOI: http://dx.doi.org/10.1017/S0305000900009818

Proffitt, J. B., Coley, J. D., & Medin, D. L. (2000). Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26 (4)*, 811 – 828. DOI: 10.10371/0278-7393.26.4.811 2000, Vol. 26, No. 4, 811-828

Quinn, P. C. (2016). Establishing cognitive organization in infancy: from perceptual grouping of objects to social classification of faces. In Balter, L., & Tamis-LeMonda, C. S. (Eds.), *Child Psychology: a handbook of contemporary issues, 3rd edition* (79 – 104). New York: Routledge.

Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, *63*, 189 – 211. DOI: http://dx.doi.org/10.1006/jecp.1996.0047

Quinn, P. C., & Eimas, P. D. (1997) A reexamination of the perceptual-to-conceptual shift in mental representations. *Review of General Psychology*, *1*, 271-287. DOI: http://psycnet.apa.org/doi/10.1037/1089-2680.1.3.271

Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: comparing adults and children. *Journal of Experimental Child Psychology*, *131*, 149 – 169. DOI: http://dx.doi.org/10.1016/j.jecp.2014.10.007

Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: the role of age and executive functioning. *PLoS ONE, 9 (1)*, 1 – 10. DOI: 10.1371/journal.pone.0085316

Rehfeldt, R. A. (2003). Establishing contextual control over generalized equivalence relations. *The Psychological Record*, *53*, 415-428. Retrieved from: https://search.proquest.com/openview/395065ad5f7385fec44fbd6281b774ec/1?pq-origsite=gscholar&cbl=41800

Rehfeldt, R. A., & Barnes-Holmes, Y. (2009). *Derived relational responding: applications for learners with autism and other developmental disabilities*. Oakland, CA: Context Press, New Harbinger Publications, Inc.

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M., & Kowalchuk, R. K. (2007). Assessing relational learning deficits in perspective-taking in children with high functioning autism spectrum disorder. *The Psychological Record, 57*, 23-47. Retrieved from: https://search.proquest.com/openview/a890a555970334d8bbb7b9ed2aaee6c2/1?pq-origsite=gscholar&cbl=41800

Rehfeldt, R. A., & Root, S. L. (2005). Establishing derived requesting skills in adults with severe developmental disabilities. *Journal of Applied Behavior Analysis*, *38 (1)*, 101 – 105. DOI: 10.1901/jaba.2005.106-03

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665-681. DOI:

https://doi.org/10.1016/S0022-5371(75)80055-7

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12,* 1 – 20. DOI: https://doi.org/10.1016/S0022-5371(73)80056-8

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. DOI: 10.1901/jeab.1997.67-275.

Roid, G. H. (2003). *Stanford-Binet intelligence scales, 5th edition*. Riverside Publishing: Itasca: IL.

Ropar, D., & Peebles, D. (2008). Sorting preference in children with autism: The dominance of concrete features. *Journal of Autism and Developmental Disorders*, *37 (2)*, 270 – 280. DOI: https://doi.org/10.1007/s10803-006-0166-2

Rosales, R., Rehfeldt, R. A., & Lovett, S. (2011). Effects of multiple exemplar training on the emergence of derived relations in preschool children learning a second language. The Analysis of Verbal Behavior, 27 (1), 61 – 74. DOI: https://doi.org/10.1007/BF03393092

Rosch, E. (1978). Principles of categorization. In E. Rosch, B. B. Lloyd, & Barbara B. B. (Eds.). *Cognition and categorization* (pp.27-48). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8,* 382-439. DOI: https://doi.org/10.1016/0010-0285(76)90013-X

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (2004). Basic objects in natural categories. In Balota, D. A., & Marsh, E. J. (Eds.). *Cognitive psychology: key readings* (448 – 471). New York: Psychology Press

Ross, B. H., & Murphy, G. L. (1999). Food for thought: cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, *38*, 495 – 553. DOI: https://doi.org/10.1006/cogp.1998.0712

Rostad, K., Yott, J., & Poulin-Dubois, D. (2012). Development of categorization in infancy: advancing forward to the animate/inanimate level. *Infant Behavior and Development*, *35*, 584 – 595. DOI: http://dx.doi.org/10.1016/j.infbeh.2012.05.005

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2 (5)*, 255 – 267. DOI: 10.1034/j.1601-183X.2003.00037.x

Ruiz, F. J., & Luciano, C. (2011). Cross-domain analogies as relating derived relations among two separate relational networks. *Journal of the Experimental Analysis of Behavior*, *95 (3)*, 369 – 385. DOI: 10.1901/jeab.2011.95-369

Ruiz-Sánchez, L. J., & Montoya-Rodríguez, M. M. (2014). A review of empirical relational frame theory articles in children and adolescents with developmental disabilities. *European Journal of child Development, Education and Psychopathology*, *2 (1)*. 17 – 26. Retrieved from: https://dialnet.unirioja.es/descarga/articulo/5761690.pdf

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. DOI: 10/1901/jeab.1988.49-95

Schwebel, A. I. & Schwebel, C. R. (1974). The relationship between performance on Piagetian tasks and impulsive responding. *Journal for Research in Mathematics Education*, *5*. 98- 104. Retrieved from http://www.jstor.org/discover/10.2307/748619?uid=3738232&uid=2&uid=4&sid=21102457621977

Shulman, C., Yirmiya, N., & Greenbaum, C. W. (1995). From categorization to classification: a comparison among individuals with autism, mental retardation, and normal development. *Journal of Abnormal Psychology*, *104*, 601 – 609. doi: http://psycnet.apa.org/doi/10.1037/0021-843X.104.4.601

Shylaja, K., & Manjula, R. (2016). Assessment of superordinate categorization in 2 – 4 years typically developing children. *Communication Disorders, Deaf Studies and Hearing Aids*, *4*, 1 – 11. DOI: http://dx.doi.org/10.4172/2375-4427.1000161

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech, Language, and Hearing Research*, *14 (1)*, 5-13. DOI: 10.1044/jshr.1401/05

Sidman, M. (1994). *Equivalence relations and behaviour: A research story*. Boston, MA: Authors Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, *74*, 127-146. DOI: 10.1901/jeab.2000.74-127.

Sidman, M. (2009). Equivalence relations and behaviour: an introductory tutorial. *The Analysis of Verbal Behavior*, *25*, 5 – 17. DOI: https://doi.org/10.1007/BF03393066

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37 (1),* 5-22. DOI: 10.1901/jeab.1982.37-5

Siegel, L. S., McCabe, A. E., Brand, J., & Matthews, J. (1978). Evidence for the understanding of class inclusion in preschool children: Linguistic factors and training effects. *Child Development*, *49*. 688–693. DOI: 10.2307/1128236

Siegler, R. S. & Svetina, M. (2006). What leads children to adopt new strategies? A microgenetic/ cross-sectional study of class inclusion. *Child Development*, *77*. 997- 1015. DOI: 10.1111/j.1467-8624.2006.00915.x

Slattery, B. & Stewart, I. (2014). Hierarchical classification as relational framing. *Journal of the Experimental Analysis of Behavior*, *101*. 61- 75. DOI: 10.1002/jeab.63

Slattery, B., Stewart, I., & O'Hora, D. (2011). Testing for transitive class containment as a feature of hierarchical classification. *Journal of the Experimental Analysis of Behavior*, *96,* 243 – 260. DOI: 10.1901/jeab.2011.96-243

Smith, C. L. (1979). Children's understanding of natural language hierarchies. *Journal of Experimental Child Psychology*, *27,* 437–458. DOI: https://doi.org/10.1016/0022-0965(79)90034-1

Smith, P. K., Smallman, R., & Rucker, D. D. (2016). Power and categorization:

power increases the number and abstractness of categories. *Social Psychology and Personality Science*, *7,* 281 – 289. DOI: 10.1177/1948550615619760

Song, L., Pruden S. M., Michnick Golinkoff, R., & Hirsh-Pasek, K. (2016). Prelinguistic foundations of verb learning: infants discriminate and categorize dynamic human actions. *Journal of Experimental Child Psychology*, *151*, 77 – 95. DOI: http://dx.doi.org/10.1016/j.jecp.2016.01.004

Soulières, I., MAttron, L., Giguère, G., & Larochelle, S. (2011). Category induction in autism: slower, perhaps different, but certainly possible. *The Quarterly Journal of Experimental Psychology*, *64 (2),* 311 – 327. DOI: 10.1080/17470218.2010.492994

Spencer, J., O'Brien, J., Riggs, K., Braddick, O., Atkinson, J., & Wattam-Bell, J. (2000). Motion processing in autism: Evidence for a dorsal stream deficiency. *Neuroreport*, *11 (12),* 2765 – 2767. Retrieved from: http://journals.lww.com/neuroreport/Abstract/2000/08210/Motion_processing_in_autism__evidence_for_a_dorsal.31.aspx

Staunton, C., & Barnes-Holmes, D. (2004). Relational frame theory, semantic priming and event related potentials: Building bridges between cognitive neuroscience and behavioural psychology. *The Irish Psychologist, 30,* 96-102.

Steele, D. L., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56,* 519-555. DOI:10.1901/jeab.1991.56-519

Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2002). Stimulus equivalence and non-arbitrary relations. *The Psychological Record, 52,* 77-88. Retrieved from: https://search.proquest.com/openview/98c1ad2857e6640a88bd13e142efe0f9/1?pq-origsite=gscholar&cbl=41800

Stewart, I., & McElwee, J. (2009). Relational responding and conditional discrimination procedures: an apparent inconsistency and clarification. *The Behavior Analyst*, *32*, 309 – 317.

Stewart, I., & Roche, B. (2013). Relational frame theory: An overview. In: Dymond, S., & Roche, B. (Eds.). *Advances in Relational Frame Theory: Research and application*. Oakland, CA: Context Press.

Stewart, I., Tarbox, J., Roche, B., & O'Hora, D. (2013). Education, intellectual development and relational frame theory. In: Dymond, S., & Roche, B. (Eds.). *Advances in relational frame theory: Research and application*. Oakland, CA: Context Press

Sundberg, M. L. (2008). *VB-MAPP: Verbal behaviour milestones assessment and placement program – protocol*. Concord, CA: AVB Press.

Tager-Flusberg, H. (1985). Basic level and superordinate level categorization by autistic, mentally retarded, and normal children. *Journal of Experimental Child Psychology*, *40 (3)*, 450 – 469. DOI: https://doi.org/10.1016/0022-0965(85)90077-3

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457-482. DOI: https://doi.org/10.1016/0010-0285(91)90016-H

Thomas, H., & Horton, J. J. (1997). Competency criteria and the class inclusion task: modelling judgments and justifications. *Developmental Psychology, 33*, 1060 – 1073. DOI: http://psycnet.apa.org/doi/10.1037/0012-1649.33.6.1060

Törneke, N. (2010). *Learning RFT: An introduction to relational frame theory and its clinical application*. Oakland: New Harbinger.

Valentin, D., & Chanquoy, L. (2012). Olfactory categorization: a developmental study. *Journal of Experimental Child Psychology*, *113*, 337 – 352. DOI: http://dx.doi.org/10.1016/j.jecp.2012.05.007

Viscaíno-Torres, R. M., Ruiz, F. J., Luciano, C., López-López, J. C., BArbero-Rubio A., & Gil, E. (2015). The effect of relational training on intelligence quotient: a case study. *Psicothema: Revista Anual de Psicología, 27*, 120 – 127. DOI: 10.7334/psicothema2014.149

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109 (1),* 163 – 167. DOI: https://doi.org/10.1016/j.cognition.2008.07.013

Vygotsky, L.S. (1962). Thought and language. Cambridge, MA: MIT Press.

Walsh, S., Horgan, J., May, R.J., Dymond, S., & Whelan, R. (2014). Facilitating relational framing in children and individuals with developmental delay using the Relational Completion Procedure. *Journal of the Experimental Analysis of Behavior*, *101,* 51-60. DOI: 10.1002/jeab.66

Watson, J., & Workman, E. A. (1981). The non-concurrent multiple baseline across individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry*, *12(3),* 257-259. DOI:10.1016/0005-7916(81)90055-0

Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence-Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for children Third Edition Manual*. New York: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale Third Edition (WAIS-III)*. San Antonio, TX: Psychological Corporation.

Weil, T. M., Hayes, S. C., & Capurro, P. (2011). Establishing a deictic relational repertoire in young children. *The Psychological Record, 61(3),* 371-390. Retrieved from: https://search.proquest.com/openview/096960a09186d12deb6e171964681d74/1?pq-origsite=gscholar&cbl=41800

Westemann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society: B, 369*, 1 – 10. DOI: http://dx.doi.org/10.1098/rstb.2012.0391

Williams, K. T. (2007). *Expressive Vocabulary Test, 2nd edition (EVT-2)*. Circle Pines, MN: Pearson Education.

Wilson-Quayle, A. & Pasnak, R. (1997). The training of class inclusion and English language skills in young Latino children. *Journal of Research in Childhood Education, 11*. 152- 162. DOI: 10.1080/02568549709594702

Winer, G. A. (1980). Class-inclusion reasoning in children: A review of the empirical literature. *Child Development, 51*. 309–328. DOI: 10.2307/1129264

Wohlwill, J. F. (1968). Responses to class-inclusion questions for verbally and pictorially presented items. *Child Development*, 449 – 465. DOI: 10/2307/1126957

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, *40 (2)*, 125 – 144. DOI: 10.1901/jeab/1988.50-125

Yamazaki, Y., Saiki, M., Inada, M., Iriki, A., & Watanabe, S. (2014). Transposition and its generalization in common marmosets. *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*, 317 – 326. DOI: http://psycnet.apa.org/doi/10.1037/xan0000027

Zentall, T. R., Galizio, M., & Critchfield, T. S. (2002). Categorization, concept learning, and behavior analysis: An introduction. *Journal of the Experimental Analysis of Behavior, 78*, 237-248. DOI: 10.1901/jeab.2002.78-237

APPENDICES

APPENDIX A

Appendix A. Trial types used in the arbitrary hierarchy phase of the second administration of the relational test (Study 1).

Property	**Statement**	**Question [Answer]**
ME (2 stimuli)	A bento is a type of animal	Is a bento a type of animal? [Yes]
		Is an animal a type of bento? [No]
	A stak is a type of animal.	Are all staks animals? [Yes]
		Are all animals staks? [No]
	A zum is a type of animal.	Are some animals zums? [Yes]
		Does the class of zums contain all of the animals? [No]
	A vink is a type of animal.	Does the class of animals contain vinks? [Yes]
		Are all animals vinks? [No]
ToF (2 stimuli)	A bento is a type of jin. Jins have lungs	Do all bentos have lungs? [Yes]
		Does the class of bentos contain members without lungs? [No]
	A stak is a type of wilk. Wilks have silver ears.	Does the class of staks contain members that have silver ears? [Yes]
		Does the class of wilks contain members without silver ears? [No]
	A zum is a type of taga. Zums have feathers.	Do all tagas have feathers? [Don’t Know]
		Do some tagas have feathers? [Yes]
	A vink is a type of ropa. Vinks have claws.	Does the class of ropas contain members with claws? [Yes]
		Do all ropas have claws? [Don’t Know]
ME (2 stimuli)	The class of animals contains bentos.	Are all bentos animals? [Yes]
		Are all animals bentos? [No]
	The class of animals contains staks.	Does the class of animals contain staks? [Yes]
		Does the class of staks contain all of the animals? [No]
	The class of animals contains zums.	Does the class of animals contain zums? [Yes]

	The class of animals contains vinks.	Are all animals zums? [No] Are some animals vinks? [Yes] Does the class of vinks contain all of the animals? [No]
ToF (2 stimuli)	The class of tagas contains zums. Tagas have beady eyes.	Does the class of zums contain members that have beady eyes? [Yes] Does the class of tagas contain members without beady eyes? [No]
	The class of ropas contains vinks. Ropas have three eyes.	Do all vinks have three eyes? [Yes] Does the class of vinks contain members without three eyes? [No]
	The class of jins contains bentos. Bentos like salty food.	Does the class of jins contain members that like salty food? [Yes] Do some jins like salty food? [Yes]
	The class of wilks contains staks. Staks eat apples.	Do all wilks eat appless? [Don't Know] Do some wilks eat apples? [Yes]
CE (3 stimuli)	A Tang is a type of jin, a jin is a type of animal.	Is a tang a type of animal? [Yes] Is an animal a type of tang? [No]
	A yalt is a type of wilk, a wilk is a type of animal.	Are all yalts animals? [Yes] Are all animals yalts? [No]
	An unda is a type of taga, a taga is a type of animal.	Are some animals undas? [Yes] Does the class unda contain all of the animals? [No]
	A sibe is a type of ropa, a ropa is a type of animal.	Does the class of animals contain sibes? [Yes] Are all animals sibes? [No]
ToF (3 stimuli)	A tang is a type of jin, a jin is a type of farro. Farros have tails.	Do all tangs have tails? [Yes] Does the class of farros contain members without tails? [No]
	A wob is a type of yalt, a yalt is a type of wilk. Wilks are fluffy.	Does the class of wobs contain members that are fluffy? [Yes]

		Does the class of wobs contain members that aren't fluffy? [No]
	A drub is a type of unda, an unda is a type of taga. Drubs are smelly.	Are all tagas smelly? [Don't Know] Are some tagas smelly? [Yes]
	A clun is a type of sibe, a sibe is a type of ropa. Cluns have blue eyes.	Does the class of ropas contain members with blue eyes? [Yes] Do all ropas have blue eyes? [Don't Know]
CE (3 stimuli)	The class of animals contains jins, the class of jins contains tangs.	Are all tangs animals? [Yes] Are all animals tangs? [No]
	The class of animals contains wilks, the class of wilks contains yalts.	Is a yalt a type of animal? [Yes] Is an animal a type of yalt? [No]
	The class of animals contains tagas, the class of tagas contains undas.	Does the class of animals contain undas? [Yes] Are all animals undas? [No]
	The class of animals contains ropas, the class of ropas contains sibes.	Are some animals sibes? [Yes] Does the class sibe contain all of the animals? [No]
ToF (3 stimuli)	The class of tagas contains undas, the class of undas contains drubs. Tagas are hairy.	Does the class of drubs contain members that are hairy? [Yes] Does the class of taga contain members that aren't hairy? [No]
	The class of ropas contains sibes, the class of sibes contains cluns. Ropas have tentacles.	Do all cluns have tentacles? [Yes] Does the class of cluns contain members without tentacles? [No]
	The class of farros contains jins, the class of jins contains tangs. Tangs eat oranges.	Does the class of farros contain members that eat oranges? [Yes] Do some farros eat oranges? [Yes]
	The class of wilks contains yalts, the class of yalts contains wobs. Wobs are purple.	Are all wilks purple? [Don't Know] Are some wilks purple? [Yes]

Note. ME = Mutual entailment; ToF = Transformation of stimulus function; CE = Combinatorial entailment.

APPENDIX B

Appendix B. Arbitrary containment assessment and training protocol.

Baseline and Generalization Assessment:

The relationship between the arbitrary stimuli is outlined in a PowerPoint presentation shown to the participant (e.g., “A blorg is inside a grap”).

1) Present trial SD (e.g., “Does a grap contain a blorg?”).
2) Provide nonspecific praise/non-contingent reinforcement for each trial (e.g., “You’re working really hard!” “I like how you’re paying attention”). Reinforce on a VR4 schedule.
3) Record responses to arbitrary containment questions per trial type, in addition to the stimulus set used.

Training Phase (Phases 1 – 4):

This procedure applies for training phases 1 – 4. For training, only Stimulus Set 1 is employed. As before, the relationship between these stimuli is outlined in a PowerPoint presentation.

At the beginning of each training session, outline how many questions the child must answer correctly in order to earn something from the stationery box (i.e., they must beat their previous score, or in the case of a first session, they must score 50% or higher).

Tokens are exchanged on an FR4 schedule so that the child can choose a sticker from the sticker box.

1) Present trial SD.
2) Provide specific praise and feedback for all correct responses (e.g., “That’s right! A grap DOES contain a blorg, because a blorg is inside a grap!”). Correct responses are also reinforced on an FR1 token schedule.
3) For incorrect responses to arbitrary containment questions, provide corrective feedback as follows:
   For “Inside” statements with “Inside” questions where “Yes” is Correct Response:
   E.g., “A blorg is inside a grap” – “Is a blorg inside a grap?”
   a) Say: “No, that’s not it” and provide corrective feedback (e.g., “A blorg *IS INSIDE* a grap, because a blorg *IS INSIDE* a grap”).
   b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

   For “Inside” statements with “Inside” questions where “No” is Correct Response:
   E.g., “A zimp is inside a crint” – “Is a crint inside a zimp”

a) Say; “No, that’s not it” and provide corrective feedback (e.g., “A crint *ISN”T INSIDE* a zimp, because a zimp *IS INSIDE* a crint!”).
b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

For “Inside” statements with “Contains” questions where “Yes” is Correct Response:

E.g., “A stak is inside a timp” – “Does a timp contain a stak?”

a) Say: “No, that’s not it” and provide corrective feedback (e.g., “A timp *DOES CONTAIN* a stak, because a stak *IS INSIDE* a timp!”).
b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

For “Inside” statements with “Contains” questions where “No” is Correct Response:

E.g., “A quig is inside a dont” – “Does a quig contain a dont?”

a) Say: “No, that’s not it” and provide corrective feedback (e.g., “A quig *DOESN’T CONTAIN* a dont because a quig is *INSIDE* a dont!”).
b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

For “Contains” statements with “Inside” questions where “Yes” is Correct Response:

E.g., “A zimp contains a crint” – “Is a crint inside a zimp?”

a) Say: “No, that’s not it” and provide corrective feedback (e.g., A crint *IS INSIDE* a zimp, because a zimp *CONTAINS* a crint!”).
b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

For “Contains” statements with “Inside” questions where “No” is Correct Response:

E.g., “A blorg contains a grap” – “Is a blorg inside a grap?”

a) Say: “No, that’s not it!” and provide corrective feedback (e.g., “A blorg *ISN’T INSIDE* a grap because a blorg *CONTAINS* a grap!”).
b) Represent the trial until the participant emits a correct response (Reinfroce these correct responses as per Step 2).

For “Contains” statements with “Contains” questions where “Yes” is Correct Response:

E.g., “A Quig contains a dont” – “Does a quig contain a dont?”

a) Say: No, that’s not it” and provide corrective feedback (e.g., “A quig *DOES CONTAIN* a dont, because a quig *CONTAINS* a dont!”).
b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

For “Contains” statements with “Contains” questions where “No” is Correct Response:

E.g., “A stak contains a timp” – “Does a timp contain a stak?”

a) Say: “No, that’s not it” and provide corrective feedback (e.g., “A timp *DOESN’T CONTAIN* a stak, because a stak *CONTAINS* a timp!”).

   b) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

4) Record first-trial (i.e., first with stimulus set) responses to arbitrary containment questions per trial type (do not record responses during correction).
5) Terminate session after a correct response has been given on each of the 16 arbitrary containment questions within the training phase.
6) If child has beaten previous score – ensure he/she chooses something from the stationery box.

Assess for generalization once student reaches 100% correct for first trial correct with arbitrary containment questions.

Generalization – Untrained Stimuli:

Using only stimulus Sets 2 – 4, follow the procedure as above for baseline or generalization (i.e., do not provide feedback). If responses do not generalize to the untrained stimuli, the participant will be reintroduced to training within that phase with a new stimulus set as per the training procedure outlined above.

Maintenance:

Follow the procedure as in baseline with:

a) Stimulus Set 1 only (i.e., trained stimulus set), and
b) Stimulus Sets 2 – 4 (i.e., untrained stimulus sets).

Arbitrary Containment

Date:	***Participant:***	***Session Number:***
Baseline	***Training***	***Generalization***	***Maintenance***
Phase:	***Assessor:***	

Mutual Entailment				
Stimulus Set	**Trial Type**	**Correct Answer**	**Response**	**Question Order**
	1	Yes		
	2	No		
	3	No		
	4	Yes		
	5	No		
	6	Yes		
	7	Yes		
	8	No		
	9	No		
	10	Yes		
	11	Yes		
	12	No		
	13	Yes		
	14	No		
	15	No		
	16	Yes		

Transformation of Function (2 Stimuli)				
Stimulus Set	**Trial Type**	**Correct Answer**	**Response**	**Question Order**
	17	Yes		
	18	No		
	19	No		
	20	Yes		
	21	No		
	22	Yes		
	23	Yes		
	24	No		
	25	No		
	26	Yes		
	27	Yes		
	28	No		
	29	Yes		
	30	No		
	31	No		
	32	Yes		

Combinatorial Entailment				
Stimulus Set	**Trial Type**	**Correct Answer**	**Response**	**Question Order**
	33	Yes		
	34	No		
	35	No		
	36	Yes		
	37	No		
	38	Yes		
	39	Yes		
	40	No		
	41	No		
	42	Yes		
	43	Yes		
	44	No		
	45	Yes		
	46	No		
	47	No		
	48	Yes		

Transformation of Function (3 Stimuli)				
Stimulus Set	**Trial Type**	**Correct Answer**	**Response**	**Question Order**
	49	Yes		
	50	No		
	51	No		
	52	Yes		
	53	No		
	54	Yes		
	55	Yes		
	56	No		
	57	No		
	58	Yes		
	59	Yes		
	60	No		
	61	Yes		
	62	No		
	63	No		
	64	Yes		

Arbitrary Containment Procedural Fidelity

Date: ____________________ **Participant:** ________________

Phase: ____________________ **Session Number:** ___________

Mutual Entailment		
Question	SD Presentation?	Appropriate Experimenter Response?
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Transformation of Function (2 Stimuli)		
Question	SD Presentation?	Appropriate Experimenter Response?
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		

Combinatorial Entailment		
Question	SD Presentation?	Appropriate Experimenter Response?
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		

Transformation of Function (3 Stimuli)		
Question	SD Presentation?	Appropriate Experimenter Response?
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		

Example of arbitrary containment questions asked.

Statement	Question	Answer
Mutual Entailment		
A blorg is inside a grap	Is a blorg inside a grap?	Yes
A zimp is inside a crint	Is a crint inside a zimp?	No
A quig is inside a dont	Does a quig contain a dont?	No
A stak is inside a timp	Does a timp contain a stak?	Yes
A blorg contains a grap	Is a blorg inside a grap?	No
A zimp contains a crint	Is a crint inside a zimp?	Yes
A quig contains a dont	Does a quig contain a dont?	Yes
A stak contains a timp	Does a timp contain a stak?	No
A grap is inside a blorg	Is a blorg inside a grap?	No
A crint is inside a zimp	Is a crint inside a zimp?	Yes
A dont is inside a quig	Does a quig contain a dont?	Yes
A timp is inside a stak	Does a timp contain a stak?	No
A grap contains a blorg	Is a blorg inside a grap?	Yes
A crint contains a zimp	Is a crint inside a zimp?	No
A dont contains a quig	Does a quig contain a don't?	No
A timp contains a stak	Does a timp contain a stak?	Yes
Transformation of Function – Mutual Entailment		
A blorg is inside a grap. A blorg is red. A grap is blue.	Is there something red inside a grap?	Yes
A zimp is inside a crint. A zimp is yellow. A crint is pink.	Is there something pink inside a zimp?	No
A quig is inside a dont. A quig is orange. A dont is silver.	Does a quig contain something silver?	No

A stak is inside a timp. A stak is black. A timp is indigo.	Does a timp contain something black?	Yes
A blorg contains a grap. A blorg is red. A grap is blue.	Is there something red inside a grap?	No
A zimp contains a crint. A zimp is yellow. A crint is pink.	Is there something pink inside a zimp?	Yes
A quig contains a dont. A quig is orange. A dont is silver.	Does a quig contain something silver?	Yes
A stak contains a timp. A stak is black. A timp is indigo.	Does a timp contain something black?	No
A grap is inside a blorg. A blorg is red. A grap is blue.	Is there something red inside a grap?	No
A crint is inside a zimp. A zimp is yellow. A crint is pink.	Is there something pink inside a zimp?	Yes
A dont is inside a quig. A quig is orange. A dont is silver.	Does a quig contain something silver?	Yes
A timp is inside a stak. A stak is black. A timp is indigo.	Does a timp contain something black?	No
A grap contains a blorg. A blorg is red. A grap is blue.	Is there something red inside a grap?	Yes
A crint contains a zimp. A zimp is yellow. A crint is pink.	Is there something pink inside a zimp?	No
A dont contains a quig. A quig is orange. A dont is silver.	Does a quig contain something silver?	No
A timp contains a stak. A stak is black. A timp is indigo.	Does a timp contain something black?	Yes

Combinatorial Entailment

A blorg is inside a grap, a grap is inside a plak.	Is a blorg inside a plak?	Yes
A zimp is inside a crint, a crint is inside a wark.	Is a wark inside a zimp?	No
A quig is inside a dont, a dont is inside a firk.	Does a quig contain a firk?	No
A stak is inside a timp, a timp is inside a vink.	Does a vink contain a stak?	Yes
A blorg contains a grap, a grap contains a plak.	Is a blorg inside a plak?	No
A zimp contains a crint, a crint contains a wark.	Is a wark inside a zimp?	Yes
A quig contains a dont, a dont contains a firk.	Does a quig contain a firk?	Yes
A stak contains a timp, a timp contains a vink.	Does a vink contain a stak?	No
A plak is inside a grap, a grap is inside a blorg.	Is a blorg inside a plak?	No
A wark is inside a crint, a crint is inside a zimp.	Is a wark inside a zimp?	Yes
A firk is inside a dont, a dont is inside a quig.	Does a quig contain a firk?	Yes
A vink is inside a timp, a timp is inside a stak.	Does a vink contain a stak?	No

A plak contains a grap, a grap contains a blorg.	Is a blorg inside a plak?	Yes
A wark contains a crint, a crint contains a zimp.	Is a wark inside a zimp?	No
A firk contains a dont, a dont contains a quig.	Does a quig contain a firk?	No
A vink contains a timp, a timp contains a stak.	Does a vink contain a stak?	Yes

<u>Transformation of Function – Combinatorial Entailment</u>

A blorg is inside a grap, a grap is inside a plak. A blorg is red. A plak is green.	Is there something red inside a plak?	Yes
A zimp is inside a crint, a crint is inside a wark. A zimp is yellow. A wark is purple.	Is there something purple inside a zimp?	No
A quig is inside a dont, a dont is inside a firk. A quig is orange. a firk is gold.	Does a quig contain something gold?	No
A stak is inside a timp, a timp is inside a vink. A stak is black. A vink is white.	Does a vink contain something black?	Yes
A blorg contains a grap, a grap contains a plak. A blorg is red. A plak is green.	Is there something red inside a plak?	No
A zimp contains a crint, a crint contains a wark. A zimp is yellow. A wark is purple.	Is there something purple inside a zimp?	Yes
A quig contains a dont, a dont contains a firk. A quig is orange. A firk is gold.	Does a quig contain something gold?	Yes
A stak contains a timp, a timp contaisn a vink. A stak is black. A vink is white.	Does a vink contain something black?	No
A plak is inside a grap, a grap is inside a blorg. A blorg is red. A plak is green.	Is there something red inside a plak?	No
A wark is inside a crint, a crint is inside a zimp. A zimp is yellow. A wark is purple.	Is there something purple inside a zimp?	Yes
A firk is inside a dont, a dont is inside a quig. A quig is orange. A firk is gold.	Does a quig contain something gold?	Yes
A vink is inside a timp, a timp is inside a stak. A stak is black. A vink is white.	Does a vink contain something black?	No
A plak contains a grap, a grap contains a blorg. A blorg is red. A plak is green.	Is there something red inside a plak?	Yes
A wark contains a crint, a crint contains a zimp. A zimp is yellow. A wark is purple.	Is there something purple inside a zimp?	No
A firk contains a don't, a don't contains a quig. A quig is orange. A firk is gold.	Does a quig contain something gold?	No
A vink contains a timp, a timp contains a stak. A stak is black. a vink is white.	Does a vink contain something black?	Yes

APPENDIX C

Appendix C. Arbitrary hierarchy assessment and training protocol.

Baseline and Generalization Assessment:

The relationship between the arbitrary stimuli is outlined in a PowerPoint presentation shown to the participant (e.g., "A fouk is a type of hib").

4) Present trial SD (e.g., "Does the class hib contain fouks?").
5) Provide nonspecific praise/non-contingent reinforcement for each trial (e.g., "You're working really hard!" "I like how you're paying attention"). Reinforce on a VR4 schedule.
6) Record responses to arbitrary hierarchy questions per trial type, in addition to the stimulus set used.

Training Phase (Phases 1 – 4):

This procedure applies for training phases 1 – 4. For training, only Stimulus Set 1 is employed. As before, the relationship between these stimuli is outlined in a PowerPoint presentation.

At the beginning of each training session, outline how many questions the child must answer correctly in order to earn something from the stationery box (i.e., they must beat their previous score, or in the case of a first session, they must score 50% or higher).

Tokens are exchanged on an FR4 schedule so that the child can choose a sticker from the sticker box.

7) Present trial SD.
8) Provide specific praise and feedback for all correct responses (e.g., "That's right! The class hib DOES contain fouks, because a fouk is a type of hib!"). Correct responses are also reinforced on an FR1 token schedule.
9) For incorrect responses to arbitrary hierarchy questions, provide corrective feedback as follows:

   For "Type of" statements with "Type of" questions where "Yes" is Correct Response:

   E.g., "A fouk is a type of hib"– "Is a fouk a type of hib?"

   c) Say: "No, that's not it" and provide corrective feedback (e.g., "A fouk *IS A TYPE OF* hib, because a fouk *IS A TYPE OF* hib").
   d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

   For "Type of" statements with "Type of" questions where "No" is Correct Response:

   E.g., "A sibe is a type of rapa"– "Is a rapa a type of sibe?"

c) Say; “No, that’s not it” and provide corrective feedback (e.g., “A rapa <u>*ISN”T A TYPE OF*</u> sibe, because a sibe <u>*IS A TYPE OF*</u> rapa!”).
d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

<u>For “Type of” statements with “Contains” questions where “Yes” is Correct Response:</u>

E.g., “A zimp is a type of yalt”– “Does the class yalt contain zimps?”

c) Say: “No, that’s not it” and provide corrective feedback (e.g., “A yalt <u>*DOES CONTAIN*</u> a zimp, because a zimp <u>*IS A TYPE OF*</u> yalt!”).
d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

<u>For “Type of” statements with “Contains” questions where “No” is Correct Response:</u>

E.g., “A A ving is a type of unda”– “Does the class ving contain undas?”

c) Say: “No, that’s not it” and provide corrective feedback (e.g., “The class ving <u>*DOESN’T CONTAIN*</u> undas because a ving <u>*IS A TYPE OF*</u> unda!”).
d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

<u>For “Contains” statements with “Type of” questions where “Yes” is Correct Response:</u>

E.g., “The class sibe contains rapas”– “Is a rapa a type of sibe?”

c) Say: “No, that’s not it” and provide corrective feedback (e.g., A rapa <u>*IS A TYPE OF*</u> sibe, because the class sibe <u>*CONTAINS*</u> rapas!”).
d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

<u>For “Contains” statements with “Type of” questions where “No” is Correct Response:</u>

E.g., “The class fouk contains hibs” – “Is a fouk a type of hib?”

c) Say: “No, that’s not it!” and provide corrective feedback (e.g., “A fouk <u>*ISN’T A TYPE OF*</u> hib because the class fouk <u>*CONTAINS*</u> hibs!”).
d) Represent the trial until the participant emits a correct response (Reinfroce these correct responses as per Step 2).

<u>For “Contains” statements with “Contains” questions where “Yes” is Correct Response:</u>

E.g., “The class ving contains undas” – “Does the class ving contain undas?”

c) Say: No, that’s not it” and provide corrective feedback (e.g., “The class ving <u>*DOES CONTAIN*</u> undas, because the class ving <u>*CONTAINS*</u> undas!”).
d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

<u>For “Contains” statements with “Contains” questions where “No” is Correct Response:</u>

E.g., “The class zimp contains yalts” – “Does the class yalt contain zimps?”

c) Say: “No, that’s not it” and provide corrective feedback (e.g., “The class yalt *DOESN’T CONTAIN* zimps, because the class zimp *CONTAINS* yalts!”).
d) Represent the trial until the participant emits a correct response (Reinforce these correct responses as per Step 2).

10) Record first-trial (i.e., first with stimulus set) responses to arbitrary hierarchy questions per trial type (do not record responses during correction).
11) Terminate session after a correct response has been given on each of the 16 arbitrary hierarchy questions within the training phase (or 12 arbitrary hierarchy questions as in Phase 2 and 4).
12) If child has beaten previous score – ensure he/she chooses something from the stationery box.

Assess for generalization once student reaches 100% correct for first trial correct with arbitrary hierarchy questions.

Generalization – Untrained Stimuli:

Using only stimulus Sets 2 – 4, follow the procedure as above for baseline or generalization (i.e., do not provide feedback). If responses do not generalize to the untrained stimuli, the participant will be reintroduced to training within that phase with a new stimulus set as per the training procedure outlined above.

Maintenance:

Follow the procedure as in baseline with:

c) Stimulus Set 1 only (i.e., trained stimulus set), and
d) Stimulus Sets 2 – 4 (i.e., untrained stimulus sets).

Arbitrary Hierarchy

Date:		Participant:		Session Number:			
Baseline		Training		Generalization		Maintenance	
Phase:				Assessor:			

Mutual Entailment				
Stimulus Set	Trial Type	Correct Answer	Response	Question Order
	1	Yes		
	2	No		
	3	No		
	4	Yes		
	5	No		
	6	Yes		
	7	Yes		
	8	No		
	9	No		
	10	Yes		
	11	Yes		
	12	No		
	13	Yes		
	14	No		
	15	No		
	16	Yes		

Transformation of Function (2 Stimuli)				
Stimulus Set	Trial Type	Correct Answer	Response	Question Order
	17	Yes		
	18	No		
	19	Yes		
	20	Yes		
	21	Yes		
	22	No		
	23	No		
	24	Yes		
	25	No		
	26	Yes		
	27	No		
	28	Yes		

Combinatorial Entailment				
Stimulus Set	Trial Type	Correct Answer	Response	Question Order
	29	Yes		
	30	No		
	31	No		
	32	Yes		
	33	No		
	34	Yes		
	35	Yes		
	36	No		
	37	No		
	38	Yes		
	39	Yes		
	40	No		
	41	Yes		
	42	No		
	43	No		
	44	Yes		

Transformation of Function (3 Stimuli)				
Stimulus Set	Trial Type	Correct Answer	Response	Question Order
	45	Yes		
	46	No		
	47	Yes		
	48	Yes		
	49	Yes		
	50	No		
	51	No		
	52	Yes		
	53	No		
	54	Yes		
	55	No		
	56	Yes		

Arbitrary Hierarchy Procedural Fidelity

Date: ____________________ **Participant:** ________________

Phase: ____________________ **Session Number:** ___________

Mutual Entailment		
Question	SD Presentation?	Appropriate Experimenter Response?
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

Transformation of Function (2 Stimuli)		
Question	SD Presentation?	Appropriate Experimenter Response?

Combinatorial Entailment		
Question	SD Presentation?	Appropriate Experimenter Response?

Transformation of Function (3 Stimuli)		
Question	SD Presentation?	Appropriate Experimenter Response?
49		
51		
52		
54		
55		
56		
57		
59		
60		
61		
63		
64		

Example of arbitrary hierarchy questions asked.

Statement	Question	Answer
Mutual Entailment		
A fouk is a type of hib.	Is a fouk a type of hib?	Yes
A sibe is a type of rapa.	Is a rapa a type of sibe?	No
A ving is a type of unda.	Does the class ving contain undas?	No
A zimp is a type of yalt.	Does the class yalt contain zimps?	Yes
The class fouk contains hibs.	Is a fouk a type of hib?	No
The class sibe contains rapas.	Is a rapa a type of sibe?	Yes
The class ving contains undas.	Does the class ving contain undas?	Yes
The class zimp contains yalts.	Does the class yalt contain zimps?	No
A hib is a type of fouk.	Is a fouk a type of hib?	No
A rapa is a type of sibe.	Is a rapa a type of sibe?	Yes
An unda is a type of ving.	Does the class ving contain undas?	Yes
A yalt is a type of zimp.	Does the class yalt contain zimps?	No
The class hib contains fouk.	Is a fouk a type of hib?	Yes
The class rapa contains sibes.	Is a rapa a type of sibe?	No
The class unda contains vings.	Does the class ving contain undas?	No
The class yalt contains zimps.	Does the class yalt contain zimps?	Yes
Transformation of Function – Mutual Entailment		
A fouk is a type of hib. Fouks have yellow tongues. Hibs have paws.	Do all fouks have paws?	Yes
A ving is a type of unda. Vings have blue noses, undas have freckles.	Do some undas have blue noses?	No
A zimp is a type of yalt. Zimps have red eyes, yalts have whiskers.	Does the class zimp contain members without whiskers?	Yes

The class sibe contains rapas. Sibes have green ears, rapas have beaks.	Do all rapas have green ears?	Yes
The class ving contains undas. Vings have blue noses, undas have freckles.	Do some vings have freckles?	Yes
The class zimp contains yalts. Zimps have red eyes. Yalts have whiskers.	Does the class yalts contain members without red eyes?	No
A hib is a type of fouk. Fouks have yellow tongues, hibs have paws.	Do all hibs have yellow tongues?	No
An unda is a type of ving, vings have blue noses, undas have freckles.	Do some vings have freckles?	Yes
A yalt is a type of zimp. Zimps have red eyes, yalts have whiskers.	Does the class yalt contain members without red eyes?	No
The class hib contains fouks. Fouks have yellow tongues, hibs have paws.	Do all fouks have paws?	Yes
The class unda contains vings. Vings have blue noses, undas have freckles.	Do some undas have blue noses?	No
The class yalt contains zimps. Zimps have red eyes, yalts have whiskers.	Does the class zimp contain members without whiskers?	Yes

Combinatorial Entailment

A fouk is a type of hib, a hib is a type of klab.	Is a fouk a type of klab?	Yes
A sibe is a type of rapa, a rapa is a type of quilb.	Is a quilb a type of sibe?	No
A ving is a type of unda, an unda is a type of taga.	Does the class ving contain tagas?	No
A zimp isa type of yalt, a yalt is a type of wilk.	Does the class wilk contain zimps?	Yes
The class fouk contains hibs, the class hib is contains klabs.	Is a fouk a type of klab?	No
The class sibe contains rapas, the class rapa contains quilbs.	Is a quilb a type of sibe?	Yes
The class ving contaisn undas, the class unda contains tagas.	Does the class ving contain tagas?	Yes
The class zimp contains yalts, the class yalt contains wilks.	Does the class wilk contain zimps?	No
A klab is a type of hib, a hib is a type of fouk.	Isa fouk a type of klab?	No
A quilb is a type of rapa, a rapa is a type of sibe.	Is a quilb a type of sibe?	Yes
A taga is a type of unda, an unda isa type of ving.	Does the class ving contain tagas?	Yes
A wilk is a type of yalt, a yalt is a type of zimp.	Does the class wilk contain zimps?	No
The class klab contains hibs, the class hib contains fouk.	Is a fouk a type of klab?	Yes
The class quilb contains rapas, the class rapa contains sibe.	Is a quilb a type of sibe?	No

The class taga contains undas, the class unda contains vings.	Does the class ving contains tagas?	No
The class wilk contains yalts, the class yalt contains zimp.	Does the class wilk contain zimps?	Yes

Transformation of Function – Combinatorial Entailment

A fouk is a type of hib, a hib is a type of klab. Fouks have yellow tongues, klabs have fur.	Do all fouks have fur?	Yes
A ving is a type of unda, an unda is a type of taga. Vings have blue noses, tagas have scales.	Do some tagas have blue noses?	No
A zimp is a type of yalt, a yalt is a type of wilk. Zimps have red eyes, wilks have feathers.	Does the class zimp contain memebrs without feathers?	Yes
The class sibe contains rapas, the class rapa contains quilbs. Sibes have green ears, quilbs have wings.	Do all quilbs have green ears?	Yes
The class ving contains undas, the class unda contains tagas. Vings have blue noses, tagas have scales.	Do some vings have scales?	Yes
The class zimp contains yalts, the class yalt contains wilks. Zimps have red eyes, wilks have feathers.	Does the class wilks contain members without red eyes?	No
A klab is a type of hib, a hib is a type of fouk. Fouks have yellow tongues, klabs have fur.	Do all klabs have yellow tongues?	No
A taga is a type of unda, an unda is a type of ving. Vings have blue noses, tagas have scales.	Do some vings have scales?	Yes
A wilk isa type of yalt, a yalt is a type of zimp. Zimps have red eyes, wilks have feathers.	Does the class wilk contain members without red eyes?	No
The class klab contains hibs, the class hib contains fouks. Fouks have yellow tongues, klabs have fur.	Do all fouks have fur?	Yes
The class taga contains undas, the class unda contains vings. Vings have blue noses, tagas have scales.	Do some tagas have blue noses?	No
The class wilk contains yalts, the class yalt contains zimps. Zimps have red eyes, wilks have feathers.	Does the class zimp contain members without feathers?	Yes

APPENDIX D

Appendix D. Class inclusion assessment and training protocol.

Baseline and Generalization Assessment:

Shuffle each set of cue cards (4 sets for stimulus selection, separated by category; 1 set for trial selection consisting of 8 class inclusion trials [one of each trial type] + 8 interspersal trials).

1) Have the student select a card for stimuli; lay out stimuli accordingly (e.g., 2 cats, 4 horses).
2) Select a card for the trial type.
3) Present trial SD.
4) Provide nonspecific praise/non-contingent reinforcement for each trial (e.g., “you’re working really hard!” “I like how you’re paying attention”); reinforce participation on schedule as determined by student behavior plan or teacher recommendation.
5) Record responses to class inclusion questions per trial type.
6) If non-class inclusion questions are answered incorrectly, make note; all non-class inclusion questions should be mastered as assessed during screening.
7) Remove the stimulus sets, rotate to a new category for stimulus selection, and repeat.
8) Terminate session after 8 class inclusion questions in total.

Intervention Phase 1:

Separate out animal pictures and stimulus selection cards for use during intervention; only animals will be used. For trial cards, use one set, containing one of each class inclusion trial type and an equal number of interspersal questions of varying trial types. Shuffle each set of cue cards (stimulus selection, trial mixing).

Place the training boxes in front of the student—one large clear plastic box for the category and two clear smaller plastic boxes for the stimulus types. State and point to the boxes: “This big box is for the category. What category do these belong to? [point to all the stimulus pictures]” [Student should say “animals”] “That’s right. Let’s write the category on the box.” [Teacher or student (depending on student preference) should write “animals” on the box]. “These two smaller boxes are for the different animals. The small boxes will go inside the animal category box.”

1) Have the student select a card for stimuli; teacher should select a trial type card for the trial. The teacher or student (depending on student preference) should write the names of the stimuli on each of the boxes (e.g. pig on one box and dog on another box), using dry erase marker.
2) Tell the student to put the stimuli in the appropriate boxes, stating the names of the stimuli, e.g. “Put the [stimulus 1/2] in the small [stimulus 1/2] box, and the [stimulus 1/2] in the small [stimulus 1/2] box”. Once the

stimuli are in the appropriate boxes, state "Great. [[stimulus 1/2]] and [stimulus 1/2] are types of animals, they belong to the animal category, so they all go inside the big animal category box." Student should place both boxes inside the larger box.

3) In the same order as the trial type card, tell the student to identify the stimulus box and the category box, e.g. if the trial type is more [stimulus 2] or more [category], then state "show me the [stimulus 2] box". Once the student correctly picks up or points to the [stimulus 2] box, state "show me the animal category box".
4) If the student makes an error in selecting the animal box (e.g. picks up the other stimulus box), demonstrate the correct response by picking up the animal box and stating "These all belong to the animal category. This is the animal category box." Restate "show me the animal category box," and repeat until student is correct.
5) Present trial SD while picking up each of the relevant boxes as you present the SD.
6) Provide specific praise and feedback for all correct responses, picking up the boxes (e.g., "That's right! There are 4 horses [pick up box]," "That's right, there are more animals [pick up box] than dogs [pick up box]!/ there are less dogs than animals!")
7) For incorrect responses to "more" class inclusion questions, provide corrective feedback as follows:
   a) Repeat the requirement to identify the stimulus and animal boxes, as in step 3/4.
   b) State, while picking up the boxes, ""That's right. [stimulus1/2] and [stimulus 1/2] are types of animals, so they all go inside the big animal category box. They all belong to the animal category but only these are [stimulus1/2], so there are more animals in the animal category box than there are [stimulus1/2] in the [stimulus1/2] box.
   c) Repeat the trial SD as in step 5.
   d) Provide specific praise and reinforcement for correct responding.
   e) Select a new set of stimuli and repeat the same trial type (specific type of "more" class inclusion) until correct first trial response with new stimulus set; go on to the next trial type.
8) For incorrect responses to "less" class inclusion questions, provide corrective feedback as follows:
   a) Repeat the requirement to identify the stimulus and animal boxes, as in step 3/4.
   b) State, while picking up the boxes, "That's right. [stimulus1/2] and [stimulus 1/2] are types of animals, so they all go inside the big animal category box. They all belong to the animal category but only these are [stimulus1/2], so there are less [stimulus 1/2] in the [stimulus 1/2] box than there are animals in the animal category box."
   c) Repeat the trial SD as in step 5.
   d) Provide specific praise and reinforcement for correct responding.
   e) Select a new set of stimuli and repeat the same trial type ("less" class inclusion) until correct first trial response with new stimulus set; go on to
9) Record first-trial (i.e., first with stimulus set) responses to class inclusion questions per trial type; do not record responses during correction.

10) If non-class inclusion questions are answered incorrectly, make note; all non-class inclusion questions should be mastered as assessed during screening.
11) Terminate session after a correct response has been given on each of the 8 class inclusion questions.

Move to next phase once student is 100% correct in selecting appropriate boxes and achieves 8/8 first trial correct with class inclusion questions.

Intervention Phase 2:

Use the same stimulus type and trial type card setup as previously, but use 3 sets of trial cards (8, 8, and 4 class inclusion questions, one of each type per set, mixed with equal numbers of interspersal questions). In this phase, verbal reference to the size of the boxes and the pre-trial requirement to identify the boxes by type/category are eliminated.

1) Have the student select a card for stimuli; teacher should select a trial type card for the trial. The teacher or student (depending on student preference) should write the names of the stimuli on each of the boxes (e.g. pig on one box and dog on another box), using dry erase marker.
2) Tell the student to put the stimuli in the appropriate boxes, stating the names of the stimuli, e.g. "Put the pigs in the pig box, and the dogs in the dog box". Once the stimuli are in the appropriate boxes, state "Now put all of them inside the animal category box." Student should place both boxes inside the larger box.
3) Present trial SD (do not pick up the boxes).
4) Provide specific praise and feedback for all correct responses, referencing and picking up the boxes and stating that they are all belong to the animal category, but only the stimulus type is the stimulus, for class inclusion trials, but without the verbal reference to the boxes (e.g., "That's right! There are 4 horses [pick up box],""That's right, there are more animals [pick up box] than dogs [pick up box]!/ there are less dogs than animals!")
5) For incorrect responses to "more" class inclusion questions, provide corrective feedback as follows:
    a) Present the SD to identify the stimulus and animal boxes, as in step 3/4 of the errorless teaching phase.
    b) State, while picking up the boxes, "they all belong to the animal category but only these are [stimulus1/2], so there are more animals than there are [stimulus1/2.
    c) Repeat the trial SD (without picking up the boxes).
    d) Provide specific praise and reinforcement for correct responding.
    e) Select a new set of stimuli and repeat the same trial type (specific type of "more" class inclusion) until correct first trial response with new stimulus set.
6) For incorrect responses to "less" class inclusion questions, provide corrective feedback as follows:
    a) Repeat the requirement to identify the stimulus and animal boxes, as in step 3/4 of the errorless teaching phase.

b) State, while picking up the boxes, "they all belong to the animal category but only these are [stimulus1/2], so there are less [stimulus 1/2] than there are animals"
c) Repeat the trial SD (without picking up the boxes).
d) Provide specific praise and reinforcement for correct responding.
e) Select a new set of stimuli and repeat the same trial type ("less" class inclusion) until correct first trial response with new stimulus set.

7) Record first-trial (i.e., first with stimulus set) responses to class inclusion questions per trial type; do not record responses during correction.
8) If non-class inclusion questions are answered incorrectly, make note; all non-class inclusion questions should be mastered as assessed during screening.
9) Terminate session after 20 class inclusion questions in total.

Move to next phase once student achieves 8/8 first trial correct and at least 19/20 correct overall with class inclusion questions.

Generalization Phase 1: Animals

Using only the animal cards, follow the procedure as above for baseline or generalization (i.e. do not use the boxes or provide feedback). If responses do not generalize without the boxes, a training phase will be implemented, following the same correction procedures as above but without boxes, just with gestures.

Generalization Phase 2: Novel categories

Follow the procedure as in baseline, with all categories represented.

Class Inclusion Baseline/Generalization Assessment

Participant: ________________________

Assessor: ________________________

Date: ________________________

☐ **Baseline** **(Session # _________)**

☐ **Generalization**

More [category] or [bigger subclass]?	+	-	+	-	+	-	+	-
More [category] or [smaller subclass]?	+	-	+	-	+	-	+	-
More [bigger subclass] or [category]?	+	-	+	-	+	-	+	-
More [smaller subclass] or [category]?	+	-	+	-	+	-	+	-
Less [category] or [bigger subclass]?	+	-	+	-	+	-	+	-
Less [category] or [smaller subclass]?	+	-	+	-	+	-	+	-
Less [bigger subclass] or [category]?	+	-	+	-	+	-	+	-
Less [smaller subclass] or [category]?	+	-	+	-	+	-	+	-

APPENDIX D

Class Inclusion/Hierarchy Training

Participant: ________________________

Assessor: ________________________

Date: ________________________

☐ **Intervention** **(Session # _________)**

Trial Type	First Trial		Correction Procedure Response						New Stimulus Set Response									
More [category] or [bigger subclass]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
More [category] or [smaller subclass]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
More [bigger subclass] or [category]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
More [smaller subclass] or [category]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Less [category] or [bigger subclass]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Less [category] or [smaller subclass]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Less [bigger subclass] or [category]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Less [smaller subclass] or [category]?	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-