Transitive inference and arbitrarily applicable comparative relations: A behaviour-analytic model of relational reasoning.

Munnelly, Anita

Transitive Inference and Arbitrarily Applicable Comparative Relations: A Behaviour-Analytic Model of Relational Reasoning

Anita Munnelly

Submitted to Swansea University in fulfilment of the requirements for the degree of Doctor of Philosophy

Swansea University

March 2013

Supervisor: Dr Simon Dymond

Summary

The transitive inference (TI) problem (i.e., if A > B and B > C, then A > C) has traditionally been considered a hallmark of logical reasoning. However, considerable debate exists regarding the psychological processes involved when individuals perform TI tasks. The current thesis therefore sought to further explore this issue with adult humans as the population sample. Following a review of the literature, the first empirical chapter, Chapter 2, adopted a traditional TI task and exposed participants to training and testing with a simultaneous discrimination paradigm. In addition, the chapter sought to examine the potential facilitative effects of awareness and repeated exposure to training and test phases on the emergence of TI. Results broadly demonstrated that awareness led to more accurate responses at test, and that for a number of participants, repeated exposure to training and test phases, allowed the targeted performances to emerge over time. Chapter 3 developed and determined the utility of a novel behaviour-analytic account of TI as a form of derived comparative relational responding. For the most part, findings revealed that the model has the potential to generate arbitrarily applicable comparative responding in adults, comparable to TI. However, findings from Chapter 3 also revealed that despite the implementation of a number of interventions, response accuracy was still weak on a number of the targeted relations. Chapter 4 developed a variant of the Relational Completion Procedure (RCP) to examine derived comparative responding to "More-than" and "Less-than" relations, as an extension of the behavioural account of TI adopted in Chapter 3. Findings revealed that, for the most part, the protocol was effective in establishing the targeted relations, and that the linearity (e.g., A<B, B<C) of training pairs was not found to effect the emergence of this pattern of responding. Chapter 5 sought to explore the transformation of discriminative functions via a 5-member relational network of "More-than" and "Less-than" relations. Findings revealed that, across four experiments, approximately half of the participants displayed the predicted patterns of performance. That is, half of the participants responded "less" to the stimuli ranked lower in the network (A and B) and "more" to the stimuli ranked higher in the network (D and E), on the basis of training with stimulus C. The utility of the current behaviour-analytic approach to the study of TI is discussed.

Declaration and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .. (candidate)

Date 26-03-13

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references.
A bibliography is appended.

Signed . (candidate)

Date 26-03-13

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed . (candidate)

Date 26-03-13

Tables of Contents

Acknowledgements

I would firstly like to say a massive thank you to Simon, for all your encouragement and support over the past number of years. I don't know how you put up with me!

I would also like to say a big thank you to Alice and Richard for being such good lab friends. A special thank you also to Claire, Louise McHugh and Andreas for your support and friendship over the years.

Also thanks to the many wonderful people from the Psychology Department that I got to enjoy many lunches and much needed nights out with. Thank you to Ting, Marcelle, Charlotte, Shakiela, JJ (you entertained me to say the least!), Paul, Ceri, Louise C, Andrew, Heidi, James, Maria, Thano and Edwin.

A special thank you also to Gary, for all your programming assistance. Also thank you to Carlo, Christoph, Jackie and Ian, for your help and assistance throughout the years.

Thanks also to my various housemates in Wales over the years: Becky, Frank, Marcelo, Isa, Rebecca, Maria, Tom and Marco. It was fun!

Thanks also to everyone from book club-Jenn, Jodie and Char.

Also a big thank you to Aine, Mairead and Denise back home. Thanks for your support over the years.

And last, but most certainly not least, thank you to my family, my sister Irene, my brother, Daniel, and my parents Bina and Jody (I'm sure your bank balance is a bit healthier now!). I couldn't have done it without your support!

Conference Presentations

Munnelly, A., & Dymond, S. (2012). Constructing Relational Sentences of More-than and Less-than. Paper presented at the European Association for Behaviour Analysis, Lisbon, Portugal, September 6th-9th.

Munnelly, A., & Dymond, S. (2011). Constructing Emergent Sentences: A Relational Completion Procedure for Training and Testing Derived Comparative Relations. Poster presented at the Annual Convention of the Association for Behaviour Analysis International, Denver, Colorado, USA, May 27th - 31st.

Munnelly, A., & Dymond, S. (2011). Relational Sentence Construction. In A. Munnelly (Chair), *Derived Relational Responding.* Paper presented at the Experimental Analysis of Behaviour Group meeting, London, 18th - 20th April.

Munnelly, A., & Dymond, S. (2010). Relational Reasoning with Derived Comparative Relations: Effects of Training and Testing Structure. In A. Munnelly (Chair), *Facilitating the Induction of Equivalence Classes and Emergence of Derived Comparative Relations.* Paper presented at the Annual Convention of the Association for Behavior Analysis International, San Antonio, Texas, USA, May 28th - June 1st.

Munnelly, A., Dymond, S., & Hinton, E. (2009). Experimental Analyses of Derived Stimulus Relations. In S. Dymond (Chair), *Relational Reasoning with Derived Comparative Relations.* Paper presented at the Experimental Analyses of Behaviour Group meeting, London, 6th - 8th April.

Publications

Munnelly, A., Freegard, G., & Dymond, S. (in press). Constructing relational sentences: Establishing arbitrarily applicable comparative responding with the Relational Completion Procedure. *The Psychological Record.*

List of Tables

Chapter 5

List of Figures

Chapter 1

Chapter 2

List of Abbreviations

AARR	Arbitrarily Applicable Relational Responding
CR	Conditioned Response
CS	Conditioned Stimulus
DMTS	Delayed Matching-to-sample
ITI	Inter-trial Interval
MET	Multiple-exemplar training
MTS	Matching-to-sample
NARB	Non-arbitrary
pREP	Precursor to the Relational Evaluation Procedure
RCP	Relational Completion Procedure
REP	Relational Evaluation Procedure
SDE	Symbolic Distance Effect
SVO	Subject-Verb-Object
TI	Transitive Inference
US	Unconditioned Stimulus
VTT	Value Transfer Theory

Chapter 1

Literature Review

Human problem solving has a long tradition of study within experimental psychology, and one way in which it has been studied is with the transitive inference (TI) problem. Transitive inference is considered a hallmark of human deductive reasoning abilities (Piaget, 1928) and describes the ability to infer a relationship between two non-adjacent stimuli following training with adjacent pairs (Dusek & Eichenbaum, 1997). To illustrate, human participants are first trained, through the provision of feedback, on a number of what are termed, adjacent, "premise" pairs, before exposure to test pairs involving non-adjacent stimuli. So, for example, a participant is presented with the premise pair AB, and is reinforced for selections of A, but not for B (A+ B-; where "+" and "–" represent the reinforced and non-reinforced stimuli, respectively). Similarly, a participant may be presented with the pair BC, where selections of B are reinforced and selections of C are not (B+C-); and also pairs CD (C+D-) and DE (D+E-). On reaching a pre-determined training criterion on all pairs, participants are presented with an inference test in the absence of feedback involving, for example, the novel B and D stimulus pair. Correct selections of B over D are then taken as evidence for TI behaviour (see Figure 1.1).

Inferential tests are not limited to the BD pairing, and following training on all of the above premise pairs (i.e., AB, BC, CD and DE), novel test trials may also include CE, AD, BE and AE. Typically however, the test trial AE is uninformative because of each stimulus' unique history of reinforcement (e.g., A) and non-reinforcement (e.g., E) during training. However, the test pair BD is a more interesting measure of TI, as both stimuli have a history of being reinforced and non-reinforced equally often during training. The ability to solve the TI task (and select B over D) is not limited to adult humans (e.g., Acuna, Sanes, & Donoghue, 2002; Ellenbogen, Hu, Payne, Titone, & Walker, 2007; Frank, Rudy, Levy, & O'Reilly, 2005; Greene, Spellman, Dusek, Eichenbaum, & Levy, 2001; Martin & Alsop, 2004; Moses, Villate, & Ryan, 2006; Van Opstal, Verguts, Orban, & Fias, 2008), and has also been widely studied among populations including young children (e.g., Bryant & Trabasso, 1971; McGonigle & Chalmers, 1984), and various non-human species (e.g., Davis, 1992; Dusek & Eichenbaum, 1997; Gillan, 1981; Lazareva, Smirnova, Zorina, & Rayevsky, 2001; Treichler & van Tilberg, 1996; Weaver, Steirn, & Zentall, 1997; Wynne, 1995, 1997; Wynne, von Fersen, & Staddon, 1992).

The *n*-term series task is the name typically ascribed to the two different types of TI tasks employed in studies with humans and non-humans. In the first task, subjects are presented with pairs of overlapping simultaneous discriminations, where one of the stimuli is reinforced (+) and the other is not (-) (see Figure 1.1). The stimuli from each discrimination partially overlap (i.e., overlapping simultaneous discriminations), and selections of stimuli B, C and D are reinforced when presented in one discrimination, and non-reinforced when presented in the other (Vasconcelos, 2008). To account for correct selections of B over D, plausible explanations lie in the realm of associative learning principles based on models of reinforcement history (e.g., Couvillon & Bitterman, 1992; Wynne, 1995, 1998), classical conditioning (e.g., Wagner & Rescorla, 1972) and value transfer (e.g., Frank, Rudy, & O'Reilly, 2003; von Fersen, Wynne, Delius, & Staddon, 1991). A second type of TI task is based on natural verbal relations, where participants may be presented with the following information regarding pairs of adjacent stimuli: Dougal is taller than Jack and Jack is taller than Ted. A test for TI then involves the following probe question: Who is taller, Dougal or Ted? By correctly answering Dougal, a subject is said to have performed a TI-like operation.

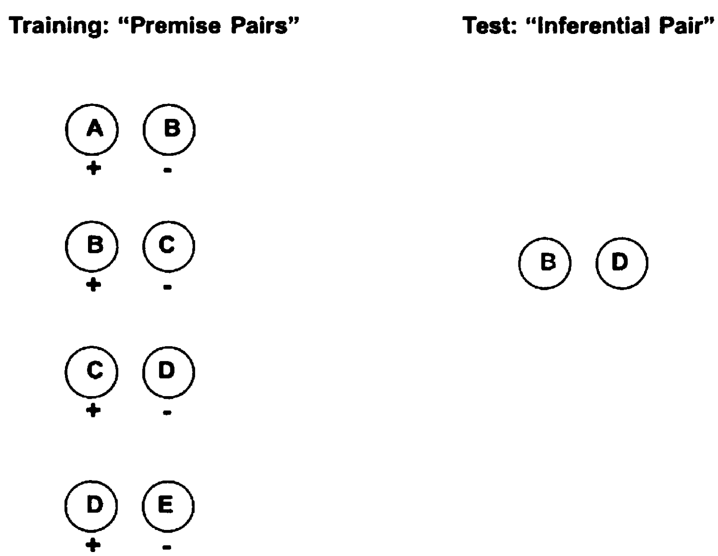


Figure 1.1. Schematic overview of a 5-term series. The column to the left contains the four overlapping simultaneous discriminations that participants are presented with during training. "+" and "-", presented under each stimulus represent the reinforced and non-reinforced stimulus in each premise pair. The column to the right contains the inferential test pair BD. The stimulus presented in red represents the predicted correct response.

A study conducted by Bryant and Trabasso (1971) has been particularly influential in the study of TI. For example, the authors examined transitive responding in young children following training and testing with a semi-verbal version (e.g., tasks that use real objects as stimuli and use language to describe the relationship between them; Wynne, 1998) of a 5-term series. Children were first trained on four premise pairs (A+B-, B+C-, C+D- and D+E-), where pairs of wooden sticks that differed from one another in terms of colour, were employed as stimuli. The children could only see the top, coloured part of the sticks, and they learnt to select the "longer" stick from each pair. Once a child had selected a stick from a pair, they were shown the actual length of the stick as a means of feedback. Thus, in effect, they were taught to associate length with colour. Testing followed, in which the BD inferential pair was presented. In this instance, B was the "shorter" stick and D was the "longer" stick. Findings demonstrated that children as young as four years old could accurately select B over D, and thus, the authors concluded that the children were capable of displaying TI behaviour.

This procedure in turn, has allowed a fully non-verbal version of the task to be developed for use with non-humans. For example, McGonigle and Chalmers (1977) exposed squirrel monkeys to training and testing with a 5-term series. However, in this study, the stimuli consisted of weighted (e.g., heavy or light) tin cans, that again, differed in terms of colour. Following training on the four premise pairs (A+B-, B+C-, C+D- and D+E-), the monkeys were exposed to a transitive test with the BD pair. Findings demonstrated that the monkeys could correctly select B over D in the BD pair, and accuracy on this test pair was significantly above chance. On the basis of these findings, variations of the task have also been developed to examine TI in hooded crows (e.g., Lazareva, Smirnova, Bagozkaja, Zorina, Rayevsky, & Wasserman, 2004; Lazareva et al., 2001), pigeons (e.g., Lazareva & Wasserman, 2012; von Fersen et al, 1991; Weaver et al., 1997), and rats (e.g., Davis, 1992; Roberts & Phelps, 1994).

As a result of both the Bryant and Trabasso (1971) and McGonigle and Chalmers (1977) studies, several versions of the task have been developed for use with human participants in the laboratory. For example, Zalesak and Heckers (2009) exposed adult participants to training and testing with a 6-term series. Participants were initially trained on five premise pairs (A+B-, B+C-, C+D-, D+E- and E+F-), which was followed by tests involving six non-adjacent pairs containing an endpoint (i.e., A or F; AD, AE, AC, BF, CF and DF), three non-adjacent pairs that did not contain an endpoint (i.e., B, C, D, or E; BD, CE and BE), and the five previously trained pairs. Results demonstrated that all participants successfully displayed transitive responding, and high accuracy was observed on all test pairs. In addition, these procedures have allowed researchers to examine the role of awareness (e.g., Greene et al., 2001; Greene, Gross, Elsinger, & Rao, 2006; Libben & Titone, 2008; Martin & Alsop, 2004; Smith & Squire, 2005), and the involvement of specific brain regions (e.g., Acuna et al., 2002; Heckers, Zalesak, Weiss, Ditman, & Titone, 2004; Koscik & Tranel, 2012; Wendelken & Bunge, 2010; Zalesak & Heckers, 2009), in the ability to solve the TI task. However, in the development of such tasks, there are a number of important factors that must be taken into consideration. For example, across both human and non-human studies, the method in which the premise pairs have been trained has varied (e.g., sequential and random), and thus, it is necessary to be aware of this difference when examining inferential performances at test. In addition, it is also important to consider whether the characteristic features of TI (e.g., symbolic distance effect) displayed by non-humans, are also seen in studies involving humans. Therefore, in order to gain a better understanding of both of these issues, it is first necessary to consider what the characteristic features are, and to consider what the different methods of training entail.

Characteristic Features of TI

As mentioned, a number of important characteristic features are also associated with performance on TI tasks involving humans (e.g., Acuna et al., 2002; Bryant & Trabasso, 1971; Colombo & Frost, 2001; Ellenbogen et al., 2007; Frank et al., 2005; Greene et al., 2001; Hinton, Dymond, von Hecker, & Evans, 2010; Zalesak & Heckers, 2009) and non-humans (e.g., Bond, Kamil, & Balda, 2003; Dusek & Eichenbaum, 1997; Gillian, 1981; McGonigle & Chalmers, 1977; Rapp, Kansky, & Eichenbaum, 1996; von Fersen et al., 1991; Wynne, 1997). These include the end-anchor effect, serial position effect and symbolic distance effect.

1. End-anchor Effect. For both humans and non-humans, accuracy is typically higher on both training and test pairs that contain one of the stimuli at the beginning or the end of the series (e.g., Smith & Squire, 2005; Wynne, 1997). Evidence for this finding stems from the fact that the terms A and E have unique histories of reinforcement (i.e., A) and non-reinforcement (i.e., E), and thus, any pair containing either of these terms should be solved more accurately than those that are devoid of them (e.g., BD; Wynne, 1998).

2. Serial Position Effect. Another common finding pertaining to performance on the training pairs is that a "U"-shaped pattern of performance is often found when accuracy is plotted and compared (e.g., Dusek & Eichenbaum, 1997; Gillan, 1981; Greene et al., 2001; Trabasso, Riley, & Wilson, 1975). This "U"-shaped pattern of performance highlights that performance is more accurate for both training pairs at the end of the series (e.g., AB and DE, in a 5-term series) in comparison to those within the series (e.g., BC and CD). However, this pattern of responding often falls foul to end-anchoring effects, in that the two training pairs that are best solved also contain the two stimuli with unique reinforcement histories, A and E (Vasconcelos, 2008). In addition, it is also predicted that performance on the training pair DE in a 5-term series should be higher than that of AB. This finding in turn is explained on the basis of associative strengths. For example, in the AB discrimination, the value of A and B (i.e., A is always reinforced and B is partially reinforced) are both positive, but in the D+E- discrimination, the value of E is always negative (e.g., never reinforced), and stimulus D has been both reinforced and non-reinforced. Thus, the relative values of the stimuli control responding, such that the value of the D+E- training pair will be greater than that of A+B- (Wynne et al., 1992).

3. Symbolic Distance Effect (SDE). A final main effect observed throughout studies of TI, is the finding that response accuracy increases and response times decrease as the number of intervening items between pairs presented at test increases, termed the symbolic distance effect (SDE; Moyer & Bayer, 1976). For example, in a 5-term series, performance accuracy should improve from BD (one intervening item) to BE (two intervening items), with the highest accuracy observed on the end-anchor pairing AE, where there are three intervening items in the test pair (e.g., Bryant & Trabasso, 1971; Frank et al., 2005; McGonigle & Chalmers, 1984; von Fersen et al., 1991; Wynne, 1997). However, in a 5-term series, this very finding may again be constrained by the end-anchor effect, in that any test pairs containing an end-anchored stimulus (i.e., A and E) will be solved more accurately than those that do not (e.g., BD; Vasconcelos, 2008). Thus, in this case, only the BD test pair when compared to BC and CD allows proper consideration of the SDE (Wynne, 1995). In order to successfully demonstrate the SDE, it is necessary to increase the number of stimuli employed to either 6 or 7 to allow for a truly transitive test of performance in which only non-adjacent test pairs that are devoid of end-anchor stimuli are analysed (see Figure 1.2).

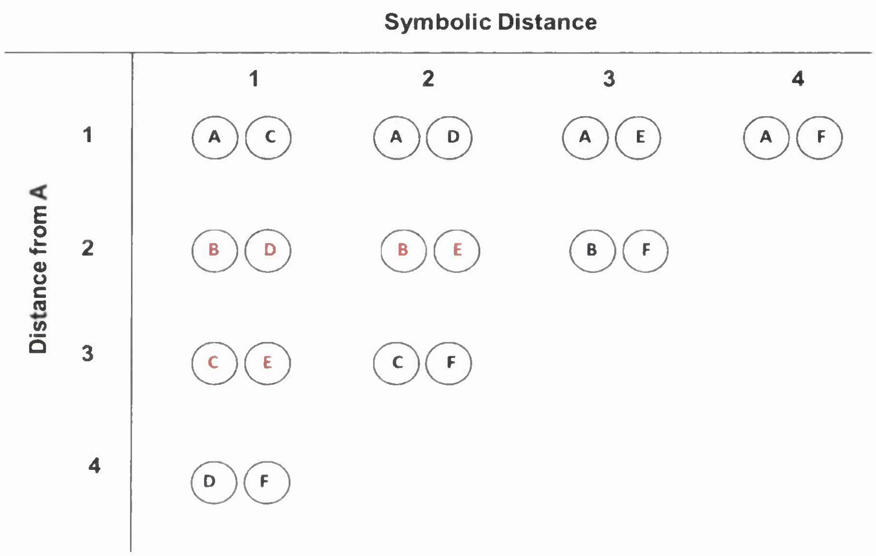


Figure 1.2. Schematic overview of the test trials presented in a 6-term series. The pairs whose letters are in red, represent the inferential test pairs.

Procedural Restrictions and Training and Testing Structures

In addition to the above characteristic features of TI, there are also some important procedural restrictions that must be considered within the experimental design to allow for the successful demonstration of transitive responding. Firstly, at least five stimuli (e.g., four training pairs) must be employed so that one of the test pairs (e.g., BD) is devoid of an end term. Also, if the number of stimuli employed within the series is increased from 5, to 6 or 7, it is possible to allow a larger number of inferential trials to be presented at test that do not contain any of the end terms (e.g., Libben & Titone, 2008; Vasconcelos, 2008; Wynne, 1998). A further benefit to extending the series is that it is possible to observe a greater and more pronounced effect for the symbolic distance effect, and to also circumvent the potential confounds of end-anchored stimuli (Vasconcelos, 2008; Wynne, 1995, 1998). Secondly, there should be no physical order to the stimuli that are presented during training. This is particularly important because at test, if the stimuli can be compared in terms of their respective heights (e.g., bigger or smaller), then the information contained within the training pairs may be irrelevant to solving the problem, which in turn could be solved without recourse to TI (e.g., Wynne, 1995).

Typically, TI studies employ one of two training schedules. The first is termed intermixed training, in which all the discriminations are trained simultaneously, in both a linear and non-linear fashion. The second method of training is termed sequential, where each discrimination is first trained by itself to criterion before the next discrimination is introduced, with training proceeding linearly (e.g., A<B, B<C, C<D and D<E; Steirn, Weaver, & Zentall, 1995; Weaver et al., 1997). Both training schedules vary considerably in the number of training trials required to reach criterion, with intermixed typically requiring a greater number of trials (Vasconcelos, 2008). It is also important to highlight the fact that both intermixed and sequential training designs permit training of the premise pairs in a linear fashion. The critical difference between the two schedules is that with sequential training, the next premise pair is not introduced until a certain criterion, typically 90% correct, or above, is achieved on the previous pair (e.g., Bernard & Giurfa, 2004; Davis, 1992; Steirn et al., 1995; Weaver et al., 1997). On the other hand, with intermixed training, the premise pairs are presented simultaneously, in both a linear and non-linear fashion, and no explicit criterion is adhered to before the next premise is introduced.

Slight variations exist among the training designs, in that the number of premise pair presentations and training blocks vary across studies. For example, some experimental designs incorporate a fixed number of training trials across all blocks (e.g., 10; Ellenbogen et al., 2007; Heckers et al., 2004), whereas others progressively decrease the number of trials presented per block as training progresses (e.g., Frank et al., 2005; Greene et al., 2001, 2006; Moses et al., 2006; Moses, Villate, Binns, Davidson, & Ryan, 2008). Another variation surrounds the number of training blocks presented. Typically, most studies present four (e.g., Frank et al., 2005; Moses et al., 2006, 2008; Smith & Squire, 2005: Exp. 3), five (e.g., Greene et al., 2001, 2006; Libben & Titone, 2008), or six (e.g., van Opstal et al., 2008) training blocks. However, in other studies, there is no fixed number of training blocks, and instead, training blocks are repeated, if necessary, until a set criterion is achieved (e.g., Ellenbogen et al., 2007; Smith & Squire, 2005).

As mentioned, intermixed training typically requires a large number of training trials to satisfy criterion, but despite this, it has been the preferred method for training the premise pairs in studies with adult humans. Typically, there are a fixed number of training blocks, in which all premise pairs are trained simultaneously. Some studies present the premise pairs out of sequence (e.g., non-linearly) with the constraint that no two neighbouring pairs (i.e., AB and BC) are presented consecutively within a training block (e.g., Ellenbogen et al., 2007; Greene et al., 2001, 2006; Smith & Squire, 2005). To illustrate, a study by Greene et al. (2001, 2006) employed four training blocks, with each consisting of 40 trial presentations. The premise pairs were presented in a non-linear fashion, in the following manner: BC, DE, AB and CD. In the first block of training, each pair was presented five times in a row, and the list repeated once (e.g., 5 x BC, 5 x DE, 5 x AB, 5 x CD, etc.). For blocks 2 and 3, the pairs were presented randomly, but the number of presentations for each trial decreased across blocks. Block 4 consisted of pairs being presented in a random order, 10 times each.

Other studies present the premise pairs in sequence (i.e., linearly), and are more prevalent throughout the TI literature than those in which the premise pairs are trained in a non-linear order. For example, in a study by Moses et al. (2006), training consisted of two phases. Phase 1 involved 10 presentations of each of the premise pairs in order (e.g., 10 x AB, 10 x BC, 10 x CD, 10 x DE and 10 x EF). Phase 2 then consisted of 5 blocks of training, each consisting of 10 trials (i.e., two presentations of each of the premise pairs), presented in a pseudo-random order. Again, across studies, the number of premise pair presentations may decrease progressively from the first block onwards, or remain constant, (e.g., Moses et al., 2006, 2008; Ryan, Moses, & Villate, 2009).

One possible criticism of intermixed training surrounds the presentation of the premise pairs in sequence. The natural progression from AB to BC and from BC to CD and so forth, may in turn lead to participants becoming aware of the underlying linear structure of the network. To address this issue, some researchers have included what are termed "distractor" trials among blocks of training in an attempt to prevent this progression from being overtly obvious. For example, Frank et al. (2005) presented stimulus combinations from other training blocks among a training block composed of mainly AB trials. In this way, there were, for example, four presentations of the AB trial with one presentation of the CD and DE trials, interspersed among this block. This same method of premise pair presentation pertained to all remaining training blocks (e.g., BC, CD, DE and EF). Thus, the "distractor" trials attempted to disrupt the descending order of the hierarchical representation of the stimuli (Frank et al., 2005).

Intermixed training also permits both "front- and back-loading" of trials. The first block of training typically involves front-loading, where there are twice as many presentations of the first three stimulus pairs at the top of the hierarchy. That is, participants may be presented with 20 instances of AB, BC and CD and only 10 instances of DE and EF (e.g., Heckers et al., 2004; Libben & Titone, 2008; Zalesak & Heckers, 2009). For back-loading which follows during the second block of training, the reverse is true. Participants are presented with 20 instances of DE and EF, and 10 instances of AB, BC and CD. The remaining training blocks are said to be "balanced", in that all premise pairs are presented for an equal number of trials. Previous findings have suggested that incorporating a block of training trials that are "front-loaded" plays a facilitative role in participants' later ability to correctly solve test trials devoid of end terms (Heckers et al., 2004; Titone, Ditman, Holzman, Eichenbaum, & Levy, 2004; Zalesak & Heckers, 2009).

Few, if any studies have incorporated a sequential training design in studies involving human participants. Most evidence for this training design has thus far emerged from studies involving non-humans (e.g., Bernard & Giurfa, 2004; Steirn et al., 1995; Weaver et al., 1997). In these studies, the AB pair is first trained by itself to criterion (typically 90% correct or above), before the next pair, BC is introduced. The pair BC is again trained by itself to criterion before the introduction of the CD pair, and training continues in this fashion for all remaining pairs of the series. In addition, sequential training also permits the direction of training to proceed in both a forward (i.e., A+B-, B+C-, C+D-, D+E-) and backward (i.e., A-B+, B-C+, C-D+, D-E+) direction, which in turn affects the resulting linear order of the series (Vasconcelos, 2008).

In addition to the different ways in which the premise pairs are trained during the TI task, variations also exist surrounding testing. Such differences centre on the amount of non-adjacent pairs presented at test, the number of times each pair is presented, and the number of testing blocks. For instance, if a 5-term series is employed, then the number of inferential test trials that may be presented is less than that of a 6-, or 7-term series. In one such study, Greene et al. (2001) employed a 5-term series, where testing involved one block, in which the novel non-adjacent test pairs, BD and AE, were presented randomly alongside the four premise pairs (AB, BC, CD and DE). Each test trial was presented eight times each, resulting in a total of 64 test trials. Similarly, Titone et al. (2004) employed a 5-term series, however, testing involved two blocks. In the first block, participants were only presented with the four premise pairs, presented 12 times each. A second test block followed, where the BD and AE test pairs were presented alongside the premise pairs. All six pairs were presented for a total of ten times each. Smith and Squire (2005) adopted a similar approach to Titone et al., with the exception that in the second block of testing, participants were exposed to only two presentations of the premise pairs, but eight presentations of the BD and AE test pairs. In a similar vein, Ellenbogen et al. (2007) also exposed participants to an initial test block in which only the premise pairs (AB, BC, CD, DE and EF) were presented, and a second test block in which the premise pairs were presented intermixed with the inferential trials. However, in this study, the authors were concerned with the effects of "offline" delays (e.g., delay between initial training and testing and delayed testing; Ellenbogen et al., 2007) on the emergence of TI behaviour, and thus, the second test block occurred either 20-minutes or 12-hours after initial training and testing.

In comparison, studies employing a 6-term series may present a greater number of non-adjacent inferential trials at test. For example, Frank et al. (2005) exposed participants to one large test block in which the five premise pairs (AB, BC, CD, DE and EF) were presented randomly, alongside four non-adjacent stimulus pairs (BD, BE, CE and AF), for a total of six times each. In addition, studies by Moses et al. (2006) and Ryan et al. (2009) involved participants being presented with an even greater number of inferential trials at test. That is, participants were presented with seven endpoint pairs (i.e., pairs that contain the two stimuli, A and F, at the end of the series; AC, AD, AE, AF, BF, CF and DF), and three non-endpoint pairs (pairs that do not contain the end stimuli A and F; BD, BE and CE) within a test block including the five premise pairs. Each test trial was presented six times, resulting in a total of 90 test trials. However, it must be noted that, in both of these studies, testing followed each block of training, and thus, participants were exposed to a total of five test blocks.

In summary, numerous studies with both humans and non-humans have identified the end-anchor effect, the serial position effect, and the symbolic distance effect as characteristic features of TI (e.g., Bond et al., 1997; Bryant & Trabasso, 1971; Greene et al., 2001; Hinton et al., 2010; McGonigle & Chalmers, 1977; Wynne, 1997). In addition, in studies examining the emergence of TI, the structure of training and test phases is also an important consideration. However, the method in which the premise pairs are presented during training (e.g., sequential or random), the number of training blocks, and the training criterion adopted varies considerably in such studies. Similar differences exist with respect to the method, and number of trials presented, during tests for inferential responding. Despite these variations, the emergence of TI behaviour was noted in all of the aforementioned studies.

The Role of Awareness and TI Performance

In addition to the characteristic features, and training and testing manipulations involved in studies examining TI, dependent measures of this behaviour often vary. Standard measures include accuracy and reaction times. However, recently, studies have begun to examine the role of awareness and its effects on inferential performances at test (e.g., Ellenbogen et al., 2007; Greene et al., 2001, 2006; Moses et al., 2006, 2008; Smith & Squire, 2005). Specifically, researchers are interested in whether or not inferential abilities are dependent on conscious awareness of the stimulus hierarchy, or whether these factors are independent of one another. Typically, the role of awareness is assessed by means of post-experimental questionnaires (e.g., Greene et al., 2001; Ryan et al., 2009; Zalesak and Heckers, 2009), where answers to the questions then form the basis of participants being classified as either aware or unaware. For example, a participant may be classified as aware if they are able to correctly identify that the stimuli can be ordered in a hierarchy, and from this ordering, it is possible to make inferential judgements (e.g., Frank et al., 2005; Moses et al., 2006, 2008; Smith & Squire, 2005). If a participant is unable to correctly identify and report either of the above, then they may be classified as unaware.

To classify participants as either aware or unaware, a numerical score is assigned on the basis of each of their responses. Awareness scores (e.g., Greene et al., 2001, 2006; Libben & Titone, 2008) typically range from 1 to 5, where a score of 5 represents definitive awareness of the stimulus hierarchy and a score of 0 represents no awareness. Most studies examining awareness have incorporated this scoring method, with a slight variation on the range of the scoring scale, and the number of questions employed existing between studies (e.g., scales ranging from 0 to 2; Moses et al., 2006, 2008; and 0 to 7; Smith & Squire, 2005). In addition to the foregoing measure of awareness, Ellenbogen et al. (2007) also incorporated a confidence rating scale, in which participants were asked to rate how confident they were on their answer, where a rating of 7 indicated the highest level of confidence.

Another variation that exists surrounding the role of awareness is whether or not participants are assigned to a "prior aware" or "prior unaware" group. Participants in the "prior" aware group are given additional information before training that the stimuli could be arranged into a hierarchy (e.g., Greene et al., 2001, 2006; Libben & Titone, 2008), and/or are told that transitive inference is a form of reasoning that allows one to make a choice between two stimuli that have not been previously been presented together (Libben & Titone, 2008). In contrast, participants in the unaware group are not given this information. Finally, a few studies have also looked at the possibility of serendipitous awareness in participants assigned to the prior unaware condition (e.g., Libben & Titone, 2008). That is, participants may become aware during the course of training or testing that the stimuli may be organised in a hierarchy, and as a result would be classified as aware.

Some researchers examining the role of awareness in TI propose that explicit awareness of the stimulus hierarchy plays an essential facilitative role in the ability to solve the task (e.g., Lazareva & Wasserman, 2010; Martin & Alsop, 2004; Moses et al., 2006, 2008; Smith & Squire, 2005). Therefore, studies have sought to compare performances on inferential trials at test for two groups of participants (e.g., Aware and Unaware) to more clearly identify the role of awareness in problem-solving tasks. In addition, these researchers propose that participants construct a linear hierarchical representation of the stimuli during training, which they can then refer to during testing to solve inferential problems (e.g., Acuna et al., 2002; de Soto, London, & Handel, 1965; Sedek & von Hecker, 2004). For example, Moses et al. (2006) exposed participants to training and testing with a 6-term series, and compared performances on novel inferential trials for two groups of participants. However, the authors did not inform participants of the relationship between the stimuli at the start of the experiment, and so awareness was assessed by means of a post-experimental questionnaire. Therefore, depending on how they scored on the awareness questionnaire, participants were either classified as "Aware" or "Unaware" of the stimulus hierarchy, and inferential performances were compared between the groups on this basis. Findings demonstrated that participants that were aware of the hierarchy displayed almost perfect responding on the premise, endpoint and non-endpoint pairs. In contrast, participants that were considered to be unaware of the hierarchy, displayed chance level performances on the non-endpoint pairs, and on some of the premise and endpoint pairs. In turn, these findings led Moses et al. (2006) to propose that awareness of the stimulus hierarchy is necessary for successful performances on the TI task (see also Lazareva & Wasserman, 2010; Martin & Alsop, 2004; Smith & Squire, 2005).

However, a number of studies have also found evidence for accurate performances on the TI task, independent of conscious awareness (e.g., Ellenbogen et al., 2007; Frank et al., 2005; Greene et al., 2001, 2006; Leo & Greene, 2008; Siemann & Delius, 1993, 1996). For instance, Greene et al. (2001) exposed two groups of participants to training and testing with a 5-term series. One group of participants (Informed) were explicitly told at the start of the experiment that the stimuli formed a hierarchy, while the other group (Uninformed) were told to learn the pairs by trial and error. In addition, participants in the Uninformed group were also required to complete a post-experimental questionnaire to assess their level of awareness during the experimental task. Results demonstrated that the Informed group displayed higher levels of accuracy on the inferential test pair BD, in comparison to participants in the Uninformed group. However, participants in the

Uninformed group performed significantly above chance on the transitive pair BD, but these measures were not correlated with post-experimental levels of awareness. This in turn led Greene et al. (2001) to propose that adult participants are capable of demonstrating TI-like behaviour in the absence of conscious awareness.

Findings from both the Moses et al. (2006) and Greene et al. (2001) studies reveal inconsistent results about whether or not conscious awareness of the underlying stimulus hierarchy is necessary for successful TI behaviour. However, the method in which awareness was measured varied across both studies. For instance, in the Greene et al. (2001) study, one group of the participants were informed of the underlying stimulus hierarchy before the experiment began, while the other group was not given this information. In addition, participants in the Uninformed group were administered a post-experimental questionnaire, to assess whether or not they became aware of the hierarchy during the task. In contrast, in the Moses et al. (2006) study, participants were not informed of any relationship between the stimuli, and participants were designated as Aware or Unaware of the stimulus hierarchy on the basis of their responses to post-experimental questionnaires. Thus, the lack of consistent, standardised measures of awareness across studies, may potentially account for the variation in results. However, one method in which researchers have recently attempted to tackle this issue is by examining the effects of pre-experimental instructions on performances at test using post-experimental measures of awareness. For example, Lazareva and Wasserman (2010) randomly assigned participants to either an Informed or Uninformed group at the start of the experiment, and additional measures of awareness were taken for both groups by means of a post-experimental questionnaire. Findings revealed that although participants in the Informed group outperformed those in the Uninformed group during tests for inferential responding, no differences were observed between the groups in terms of post-experimental measures of awareness. That is, at the end of the experimental task, participants in the Informed group were no more aware as to the nature of the experiment than those in the Uninformed group Thus, Lazareva and Wasserman (2010) found evidence suggesting that pre-experimental instructions do not enhance awareness of the experimental task. These findings in turn are important and question whether awareness is central to the ability to solve the TI task.

Understanding the role of awareness in tasks such as the aforementioned is important for a number of reasons. For example, if solving problems such as the TI task is dependent on conscious awareness, then this would suggest that humans employ only one learning strategy to solve such tasks. On the other hand, if, a dissociation between learning, and awareness exists, then this suggests that humans may employ a number of strategies to solve the task (e.g., Lovibond & Shanks, 2002). Answers to these questions are also important in that they may help to determine whether findings from the TI literature involving non-human participants, can be generalised to humans (e.g., Lovibond & Shanks, 2002; Moses et al., 2006). In order to tackle this issue, a vast amount of literature has amassed examining the role of awareness in Pavlovian (respondent) conditioning (for a review, see Lovibond & Shanks, 2002). For example, in a Pavlovian conditioning procedure, a neutral stimulus (conditioned stimulus, CS) is paired with an unconditioned stimulus (US; e.g., electric shock), which as a consequence, comes to evoke a conditioned response (CR). The CR is then taken as evidence that an association between the CS and US has been learned.

An important issue that researchers examining the emergence of respondent conditioning have sought to clarify is whether this pattern of responding is established independently of awareness. In a detailed review of the literature on respondent conditioning and awareness in humans, Lovibond and Shanks (2002) propose that the method in which awareness is currently assessed (e.g., post-experimentally), may account for the inconsistency in results. Thus, Lovibond and Shanks (2002) propose that the optimal method in which awareness may be measured, involves taking concurrent measures of either US expectancy (during CS presentations) or CS-US contingency (between trials). In this instance, *US expectancy*, refers to the fact that after the presentation of a particular CS, participants may become aware of the imminent delivery of a specific US, whereas *contingency awareness* refers to the knowledge that a specific CS predicts a specific US. Furthermore, Dawson and Reardon (1973) state that there is strong evidence showing that post-experimental measures of awareness underestimate contingency awareness. Therefore, if such methods were employed during TI tasks, in which participants were asked to report their reasoning for selecting a certain stimulus from a pair, this may help to provide a more reliable measure of awareness.

In summary, there still is considerable debate regarding whether the ability to solve the TI task is dependent or independent of conscious awareness. In addition, the method in which awareness is measured varies considerably. A potential solution to this issue was proposed by Lazareva and Wasserman (2010), in which the authors examined the effects of pre-experimental instructions on performances throughout the task, by use of post-experimental measures. However, following a review of the methods in which awareness is examined during respondent conditioning tasks, Lovibond and Shanks (2002) have proposed that perhaps a more effective method involves taking concurrent measures during the experimental task. If such methods were incorporated during TI tasks, then this may help to more clearly identify the factors controlling the emergence of this behaviour. However, before such measures can be explored, it is necessary to first gain a clearer understanding of the proposed accounts and theories that attempt to explain TI.

Models and Theories of Transitive Inference

To account for the demonstration of, and many of the characteristic effects associated with TI in human participants, both cognitive and behavioural theories have been proposed thus far. Cognitive theories have focused on the issue of whether or not participants solve the TI task by applying formal logic to the previously learnt premise pairs (e.g., Braine, 1998; Byrne & Johnson-Laird, 1989; O'Brien, 1998), or whether the independently learned premises are progressively integrated into a unified mental representation, supposedly spatial in nature (e.g., Acuna et al., 2002; de Soto et al., 1965; Huttenlocher, 1968; Trabasso & Riley, 1975). However, the finding that children as young as four years old have demonstrated successful performances on a TI task (e.g., McGonigle & Chalmers, 1984; Trabasso et al., 1975), and the suggestion that conscious awareness may not be necessary for successful performance, has led some researchers to propose that perhaps simpler behavioural theories based on the principles of associative learning may be sufficient accounts of this behaviour in humans (e.g., Couvillon & Bitterman, 1992; Frank et al., 2003; von Fersen et al., 1991; Wagner & Rescorla, 1972; Wynne, 1995, 1998).

Cognitive Theories

Some of the earliest accounts of TI in humans emerged from the cognitive literature on formal rules of logic (e.g., Braine, 1998; Hunter, 1957). However, support for the application of logical rules in solving the task has wavered, leaving theories of internal mental models (Byrne & Johnson-Laird, 1989; Johnson-Laird & Byrne, 1991), and linguistic theory (e.g., Clark, 1969) to dominate the field. In saying this, however, there are three main alternative theories dominating research efforts on this topic, the first of which is based on the rules of formal logic.

Hunter's Operational Model (1957). Hunter's (1957) Operational Model applies the rules of formal logic to the TI task. In its simplest form, performance is best when the premises are trained in a linear order, where the information contained within the premises progresses naturally from one to the next (Hunter, 1957). For example, when given the following information:

A is larger than B

B is larger than C

the premises progress naturally from the first to the second. In order to solve the problem "A is larger than C", Hunter (1957) proposes that one simply deletes the middle term "B" to draw a valid conclusion.

However, in some instances, the information contained within the premises does not follow this linear order, and in such cases, it may be necessary to reduce the information into a somewhat simpler form. So, for example, if a problem was presented with the following information contained within the premises:

B is smaller than A

B is larger than C

two cognitive operations need to be performed. The first involves *converting* the second premise to "C is smaller than B". The second operation then involves *re-ordering* the two premises into "C is smaller than B" and "B is smaller than A", to allow a valid conclusion to be drawn. In a study involving a combination of problems with varying levels of difficulty, Hunter (1957) found that it took children longer to solve the problems that involved conversion and re-ordering than those involving neither operation, and that conversion is a somewhat simpler operation than re-ordering.

The Image Theory. One of the most popular cognitive theories of TI has focused on ordered mental representations of the basic premise information (de Soto et al., 1965; Huttenlocher, 1968; Trabasso 1975, 1977; Trabasso & Riley, 1975). The theory postulates that humans construct a metaphorical mental line that contains each of the items encountered in training in their appropriate "spatial" order (Johnson-Laird, 1972). The fundamental ability to make a subsequent relational judgment is thought to rely heavily on the abstraction of information from this internal mental line (Johnson-Laird, 1972). There are two main proposals as to how and when this mental line is formed. The first, proposed by Bryant and Trabasso (1971) argues that training establishes a separate representation of each premise, and it is only during testing that this information becomes integrated into the mental line. The second, and most popular proposal, argues that all premises are integrated into the mental line during training, and performance at test is governed by the spatial representation of this information in the mental line (de Soto et al., 1965; Huttenlocher, 1968).

To explain the method by which the mental line is constructed, de Soto et al. (1965) argue that the comparatives, such as better and worse, even though they are not spatial in nature, refer to different ends of the line or scale. The term "better" is considered to represent the "good" end of the scale, and the term "worse" represents the "bad" end of the scale. As humans have a natural tendency for a preferred direction of working when constructing arrays (Evans, Newstead, & Byrne, 1993), the comparative terms will be inserted into the array, from "best" to "worst", in either a top-down fashion for a vertical array, or from left-to-right in a horizontal array. That is, the end items of the line are identified first, and stimulus A, the always reinforced stimulus (i.e., the best), is associated with one end of the array, and stimulus E (i.e., the worst), the never reinforced stimulus is associated with the opposite end. As training continues, the other stimuli are progressively incorporated into their appropriate positions on the line. Selections at test then involve a spatial search along the line to locate the correct answer. To account for the characteristic

SDE, the model proposes that the further apart the two test items are in the array, the easier it will be to locate one of them along a search of the array (i.e., larger differences are easier to detect than smaller ones; Acuna et al., 2002).

The Linguistic Theory. The most recently developed and final theory proposed to account for TI performance is based on the linguistic theory developed by Clark (1969). The central tenet to this model is that the process of deduction when engaging in a TI task is parallel to the processes involved in everyday comprehension. Three main principles are incorporated into the theory. The first includes the *principle of lexical marking*, which proposes that certain lexically marked relations are harder to understand and remember than others. That is, terms such as "better" and "worse" are considered unmarked and marked comparatives respectively, and have different meanings. For instance, imagine a journalist reports: "Roger Federer is better than Andy Murray". In this instance, the term "better" is an unmarked comparative and describes the relative goodness of two entities, but the use of the term "better" does not inform the reader of the absolute goodness of both players (i.e., whether they were both good or bad). It may be that both players performed well, or both players performed badly. However, if the journalist reported, "Roger Federer is worse than Andy Murray", then the reader has information regarding both player's relative goodness in the game, and absolute information about their playing (i.e., that they both played badly; Evans et al., 1993). To further illustrate the difference between the two types of comparatives, consider an example of the word "horse" used in everyday language. The word horse is considered an "unmarked" comparative, as it could refer to either a male or female horse. On the other hand, the word mare would be considered "marked", as it can only refer to females. Thus, the unmarked form is the basic form, while the marked form contains additional material. When this principle is applied to the TI task, it should be easier to solve unmarked than marked comparatives, because they lack the additional information contained within an unmarked comparative (Johnson-Laird, 1972). For example, in the following problems, it should be easier to solve:

A is better than B

B is better than C

than

C is worse than B

B is worse than A

because the first problem contains the unmarked comparative "better", and so an individual only receives information regarding A and B. However, the second problem contains the marked comparative "bad", and requires the individual to store more information and compare the relative degrees of badness in order to conclude an answer (Johnson-Laird, 1972).

A second principle of the linguistic theory is termed the *primacy of functional relations.* The basic assumption of this principle is that when an individual is presented with the following problem: "Green is smaller than blue", green is encoded as small + (or smaller), and blue as small. If, in a TI task, an individual was presented with the following premise:

A is worse than B,

Clark (1969) proposes that the individual understands that A and B are both bad quicker than they can comprehend their relative degrees of badness, and thus, the resulting underlying representation would be: (A is bad) more than (B is bad).

The third and final principle of the theory surrounds the issue of *congruency* between the information contained in the premises and the form of the question. For example, if participants are presented with "A is worse than B", it is assumed that this information is stored as A is "more bad" and B is "less bad" (Johnson-Laird, 1972). If the following probe question was of the form "Which is better?", then the information between the premises and the question would be incongruent with the representation in memory. Thus, it would be necessary to convert either the information contained within the premises, or the form of the question to one that is congruent with the question.

Summing up the Cognitive Theories

Each of the three cognitive theories, make independent assumptions as to how the TI task is solved. Hunter's (1957) Operational Model focuses on the natural order of the premises, and the cognitive operations of conversion and re-ordering, and is perhaps the least popular of the three theories today. Despite this, all three theories are in agreement that the congruency between the nature of the premises and the question posed plays a vital role in solving the TI task, but only Clark's (1969) linguistic theory attempts to incorporate an explanation for this into his model. Clark's Linguistic Theory is concerned with the causes of difficulties in solving TI problems, whereas de Soto et al.'s (1965) Image Theory focuses on the mental processes that are needed to make transitive inferences. Most researchers tend to agree that participants construct arrays and mental images of the information in their mind (e.g., Breslow, 1981; Potts, 1974; Riley & Trabasso, 1974; Trabasso & Riley, 1975), but the Image theory poses serious doubts regarding the validity of these accounts of TI. For example, it is unclear how a mental line is formed, and when proposed as a strategy used to solve the TI task, it provides little, if any explanatory value (see Vasconcelos, 2008). A failure to specify the underlying processes involved in the formation of the mental line also raises the question as to why the construction of the line is dependent on the end items of the series being located first. As a result, it is unclear how the theory would cope with different training designs, such as sequential training. For example, the end item, E (in a 5-term series) is the final stimulus to be encountered during training, and thus, would only be integrated into the line at the very end, which is contrary to the proposal that the end items are always located first (Vasconcelos, 2008).

Furthermore, the failure of cognitive theories to adapt their training and testing designs, if and when transitive behaviour breaks down, is also worth considering. However, this may well be explained on the basis of the underlying philosophical approach taken by cognitive psychology. For example, cognitive psychologists propose that individuals adopt meditational processes (i.e., the processing of information) as a means of describing the phenomena of interest. Thus, cognitive researchers are interested in the mental act or process by which individuals acquire information. On the other hand, behaviour-analytic researchers propose that behaviour such as that observed in the TI problem, can be explained without reference to mental events, and thus, the conditions that give rise to this behaviour, and the ways in which it can be manipulated are central to its predictions. Indeed, a number of behavioural theories of TI have been incorporated in studies involving both humans and non-humans, which will be addressed in the next section.

Behavioural Theories

Value Transfer Theory. One prominent behavioural account of performance on the TI task emerged from studies involving non-humans. Initially proposed by von Fersen et al. (1991), Value Transfer Theory (VTT), based on the principles of associative learning, attempts to provide a behavioural account of TI performance in non-humans, and has since, led to the development of a computational model of TI for use with humans (Frank et al., 2003). According to VTT, associative value transfers from the reinforced stimulus (S+) to the non-reinforced stimulus (S-) in a simultaneous discrimination. Thus, stimuli gain their value through their previous association with reward and non-reward. However, in addition, each stimulus also gains some value from its previous pairings with other rewarded stimuli (Vasconcelos, 2008). To explain how value transfers, and differential values accrue across the stimuli, VTT proposes that stimulus A has the strongest reinforcement history, and therefore, transfers part of its value to stimulus B, also weakly affecting stimulus C. On the other hand, stimulus E has a negative reinforcement history, which affects stimulus D, and again weakly affects stimulus C. So by the end of training, each stimulus will end up with a different value (e.g., A = + 2, B = + 1, C = 0, D = - 1, E = - 2). Correct selections of one stimulus over another at test can be accounted for by the differential values that stimuli amass during training (e.g., B = + 1 and D = - 1; Libben & Titone, 2008). Thus, selections of B over D at test are controlled by the difference in values between the two stimuli, and of which favours stimulus B (Wynne, 1995).

Most of the characteristic features of TI are captured by this account. For example, the account has predicted and found that test pairs containing one of the end anchor stimuli are solved better than those that do not, and also that performance on the training pairs AB and DE is typically more accurate than on the middle training pairs BC and CD (Van Elzakker, O'Reilly, & Rudy, 2003). Similarly, VTT is also able to account for the SDE, in that the theory proposes that the stimuli with greater associative strength differences result in a stronger choice preference for the more positive of the two test stimuli (Frank et al., 2005).

However, the disadvantages of this theory become evident when different procedures for training the premise pairs are employed. As VTT was initially proposed in the context of intermixed training, its predictions are questionable when applied to a sequential training schedule (Vasconcelos, 2008). With sequential training, each premise pair is first trained by itself to criterion before the next discrimination is introduced. On the basis of this method of premise pair presentation, stimulus B would receive a value of 1, when presented in the A+B- discrimination, and a value of 2, when presented in the B+C- pairing. Similarly, stimulus D would receive a value of 1 from the C+D- discrimination, and a value of 2 from the D+E- discrimination. So, by the end of training, stimuli B and D each end up with a value of 3, making it more difficult for the model to predict stimulus selection on the BD transitive probe pair (Vasconcelos, 2008). In addition, the values of stimuli B and D are greater than the values of the end-anchor stimuli A and E (e.g., 2).

Reinforcement-based Theories. In an attempt to tackle the problems associated with VTT, a number of reinforcement-based theories have been developed to account for TI performance in non-humans. All theories differ in their level of complexity and ability to account for all of the characteristic features (e.g., SDE). On the other hand, the theories are similar in the sense that they rely on reinforcement-driven mechanisms that examine performance on a trial-by-trial basis (Wynne, 1995). The theories have emerged chronologically, where the failure of one theory to account for TI performance, following a specific training procedure (i.e., intermixed or sequential), has led to the development of another model.

Bush-Mosteller (1955). The first and simplest is the Bush-Mosteller theory (Bush & Mosteller, 1955), which was initially proposed by Couvillon and Bitterman (1992). The basic premise of this theory is that reward will increase the probability of a response, whereas non-reward will decrease a subsequent response. The theory also makes the following three assumptions:

1. Each stimulus acquires its own unique value.
2. This value changes only after a choice of that stimulus.
3. The relative value of this stimulus then governs choice on each trial.

Thus, the value of a stimulus is incremented following each of its presentations according to the following parameters:

$$V(X)_{i+1} = V(X)_i + U\beta * (1 - V(X)_i) \text{ (for rewarded stimuli)}$$
$$V(X)_{i+1} = V(X)_i - D\beta * V(X)_i \text{ (for unrewarded stimuli)}$$

In this instance, $V(X)_i$ is the value of stimulus X on the trial *i*, $U\beta$ is the rate parameter determining the effect of a reward, and $D\beta$ is the rate parameter for the effect of non-reward (Wynne, 1995).

Luce's choice rule (1959) is then incorporated to account for the subsequent choice of a stimulus within a certain pair:

$$(p(X|XY) = V(X)/(V(X) + V(Y)))$$

So, for example, every time the training pair A+B- is presented and A is selected, the value of stimulus A is increased by a small amount. In contrast, every time stimulus B is chosen, there is a small decrease in its reinforcement value. Thus, after a number of exposures to this particular training trial, choice of A will govern responding as the higher value of A will outweigh that of B. This pattern of responding forms the basis for responses on all other training pairs, so that by the end of training, the stimulus values will be ranked in order of the implied series (e.g., A>B>C>D>E; Wynne, 1998).

Using pigeon data obtained by von Fersen et al. (1991), the ability of the theory to predict performance on training and test pairs in a 5-term series was tested. The theory correctly captured the signature effects of TI including the U-shaped pattern of performance on training pairs, the end-anchor effect, and the SDE. To account for correct selections of B over D on the BD test trial, the theory found that the relative value of B is greater than that of D, simply because animals tend to make more errors on the C+D- pair over the B+C- training pair (Couvillon & Bitterman, 1992; Wynne et al., 1992).

Despite the success of the theory in accounting for the characteristic effects, there are some limitations to its predictions. For example, when the series is extended to a 7-term series, by the addition of a stimulus at each end of the series (e.g., X+A- and E+F-), problems arise when the series is "closed" (e.g., training the first and last terms of the series together; von Fersen et al., 1991). That is, when the previously rewarded and non-rewarded stimuli (A and F) are trained together, and their reward/non-reward ratios reversed (e.g., A+F-), the model under-predicts performance on all training pairs, and performance on all trials at test fell to chance levels (Wynne, 1998). Also, the theory runs into difficulty when the order in which the training pairs are presented changes. As the theory operates on a trial-by-trial basis, it is capable of predicting correct transitive responding when the pairs are presented in a forward fashion (i.e., A+B-, B+C-, C+D- and D+E-). However, the same does not apply when the premise pairs are trained in the reverse order (i.e., in a backward fashion: D+E-, C+D-, B+A- and A+B-), and the theory has serious difficulties predicting choice on the test trial BD (Wynne, 1998).

Rescorla-Wagner (1972). In an attempt to tackle the potential drawbacks associated with the Bush-Mosteller (1955) theory, Rescorla and Wagner (1972) proposed a modified version of this theory based on the principles of classical conditioning. The new theory rejects Bush and Mosteller's assumption of unique or independent values for each stimulus, and thus, loses a parameter, because now, the differential effects of reward and non-reward have no impact on performance (Wynne, 1995). Instead, the Rescorla-Wagner theory assumes that stimuli compete for a limited amount of associative strength (Wynne, 1998). To illustrate, say, stimuli A and B are presented during training, and stimulus A is reinforced and B is not. If stimulus A is correctly chosen, then the value of A should be updated according to the following rule:

$$V(X)_{i+1} = V(X)_i + \beta * (1 - [V(X)_i + V(Z)_i])$$

In this instance, X is the reinforced stimulus, Y is the non-reinforced stimulus, $V(X)_{i+1}$ is the updated value of stimulus X following the current trial, $V(X)_i$ is the value of stimulus X before the current trial, β is a learning parameter, and $V(Z)_i$ is the context in which the stimuli are encountered and any commonalities between the two stimuli presented. The reinforcement-based asymptote = 1 (Vasconcelos, 2008).

If stimulus B is incorrectly chosen, then its value will be updated according to the following rule:

$$V(Y)_{i+1} = V(Y)_i - \beta * [V(Y)_i + V(Z)_i]$$

Hence, the stimuli's values are updated on a continuous basis, and again using Luce's choice rule, the stimulus with the most value will be chosen (Vasconcelos, 2008). The following scaling parameter (α) is also added to account for response probabilities:

$$P(X|XY) = \frac{1}{1 + \varepsilon^{-\alpha(2r-1)}}$$

In this instance, $P(X|XY)$ is the probability of choosing X when presented alongside Y (Vasconcelos, 2008).

Again, using von Fersen et al.'s (1991) pigeon data, the theory can account for the end-point effect, serial position effect, symbolic distance effect and predicted BD performance in a 5-term series. The main advantage that the theory holds over Bush and Mosteller's is that it can now account for training in both a forward and backward series. However, the different values that the model assigns to each stimulus, limits its ability to account for performance when the series is extended to a 7-term closed, circular series (Wynne, 1998).

Wynne's Configural Model. In an attempt to overcome the remaining problems associated with a 7-term closed circular series, Wynne (1995, 1998) proposed a modified version of the Rescorla-Wagner theory, which includes an additional assumption. Wynne (1995, 1998) proposed that the value of a stimulus is bound to the context in which it is presented. So, for example, the value of stimulus B in the context of a B+C- trial is distinct from the value it obtains in the context of A+B-. Thus, every time a stimulus is presented in a new context, it is considered a new stimulus, and therefore, has a different value (Wynne, 1995). The equations are again modified, and the new configural value [$V(X|XY)$ and $V(Y|XY)$] is added to the previous equations, by the parameter, y:

$$V(X|XY)_{i+1} = V(X|XY)_i + \beta * [1 - V(X|XY)_i]$$

$$V(Y|XY)_{i+1} = V(Y|XY)_i - \beta * V(Y|XY)_i$$

The stimulus-updating rule remains the same, and Luce's choice rule is modified to account for the configural presentations:

$$r = \frac{V(X) + V(2) + {}_yV(X|XY)}{V(X) + V(Y) + 2V(Z) + {}_y[V(X|XY) + V(Y|XY)]}$$

By combining elemental and configural stimulus values, the theory is able to overcome the limitations of the other two models, and correctly predict all of the signature effects, as well as correct transitive choice in a 7-term closed circular series.

Siemann-Delius Model. Siemann and Delius (1998) developed a further theory to account for transitive responding in animals, which is based on Luce's (1959) learning operator. According to Luce's model, values in a simultaneous discrimination (e.g., X+Y-) update on a trial-by-trial basis, according to the following equations:

$$V(X)_{i+1} = V(X)_i + \beta_+ V(X)_i \text{ (for rewarded stimuli)}$$

$$V(Y)_{i+1} = V(Y)_i - \beta_- V(Y)_i \text{ (for unrewarded stimuli)}$$

In this instance, $\beta+$ is the learning parameter, which corresponds to reinforcement, while $\beta-$ corresponds to non-reinforcement. A further equation (see below) accounts for the probability of an animal choosing X, in a given X+Y- discrimination.

$$P(X|XY) = \frac{V(X)}{V(X) + V(Y)}$$

However, like the other theories, Luce's also encounters some difficulties in accounting for transitive behaviour. For example, the theory is unable to account for an animal's failure to behave transitively even though they have successfully learned the premise pairs. In an attempt to overcome these difficulties, Siemann and Delius (1998) included a εk modification to the theory. With this modification, and similar to Wynne's configural approach, each stimulus has an elemental and configural value, which is updated on a trial-by-trial basis during training. The elemental values are updated according to the following equations, where ε is a parameter that represents the weight of the changes to the elemental values:

$$V(X)_{i+1} = V(X)_i + \beta_+ V_i P(X|XY)\varepsilon \text{ (for rewarded stimuli)}$$
$$V(Y)_{i+1} = V(Y)_i - \beta - V(Y)_i P(Y|XY)\varepsilon \text{ (for unrewarded stimuli)}$$

The configural values are also updated according to the following equations, where $_k = 1 - \varepsilon$

$$V(X|XY)_{i+1} = V(X|XY)_i + \beta_+ V(X|XY)_i P(X|XY)_K \text{ (for rewarded stimuli)}$$
$$V(Y|XY)_{i+1} = V(Y|XY)_i - \beta - V(Y|XY)_i P(Y|XY)_K \text{ (for unrewarded stimuli)}$$

As both the elemental and configural values are updated on a trial-by-trial basis, the probability of choice on these trials is seen in the following equation:

$$\frac{P(X|XY) = \quad V(X) * V(X|XY)}{V(X) * V(X|XY) + V(Y) * V(Y|XY)}$$

In the Siemann and Delius (1998) theory, the parameter ε is important, as when $\varepsilon = 0$, successful TI performance is predicted. However, as the value of ε increases so do the predictions for successful TI behaviour. In addition, the current theory can account for the serial position effect and the SDE.

Summing up the Behavioural Theories

Von Fersen et al.'s (1991) Value Transfer Theory (VTT) argues that choice during inferential probe trials is governed by the value that each stimulus has accrued during training which is based on reward and non-reward experienced.

Similar to these predictions, Bush-Mosteller (1955), and Wagner-Rescorla's (1972) reinforcement-based theories argue that the relative values of each stimulus, governs choice during training and testing. However, the method by which a stimulus gains its value differs between VTT and the reinforcement-based theories. According to VTT, say, stimulus B gains its value through a direct transfer of value from, for example, the always reinforced stimulus A, on the A+B- trial, and a certain amount when rewarded on the B+C- trial (Wynne, 1995). On the other hand, the reinforcement theories propose an indirect transfer of value from a reinforced stimulus to a non-reinforced stimulus. That is, stimulus A, which is always reinforced, gains value faster than stimulus B, which is only partially reinforced because, as stimulus A gains value, it protects stimulus B from any further loss of value that would occur if it were incorrectly chosen on any remaining A+B- trials, and thus, any future selections on this trial will be allocated to stimulus A (Wynne, 1995).

All of the aforementioned theories are able to correctly predict performance on the training pairs, the characteristic effects of TI, and performance on the inferential test pair BD. However, all have encountered their own problems. For example, VTT is unable to account for performance at test following sequential training. Similarly, the Bush-Mosteller theory is unable to predict performance when the pairs are trained in a backward fashion. Finally, the reinforcement-based theories cannot predict performance in a 7-term closed, circular series. These shortcomings in turn have been addressed by Wynne's (1995, 1998) configural theory and Siemann and Delius' (1998) theory, which overall, provide the best fit for all of the characteristic performance features associated with TI. For example, Wynne's theory accurately predicts performance following both intermixed and sequential training, and following training that proceeds in, either a forward or backward fashion. In addition, the theory can account for all of the signature effects, and accurately predicts performance on a closed circular series. The Siemann-Delius theory can account for successful training and transitive performances for human participants exposed to a 6-term series. In addition, the theory can account for above-chance performances in a circular series (Vasconcelos, 2008).

However, it must be considered that on the basis of findings from research on stimulus equivalence and derived stimulus relations, human behaviour can come under complex control. For example, when verbally sophisticated humans are trained on a number of conditional discriminations, untrained relations often emerge, in the absence of explicit reinforcement (e.g., Hayes, Barnes-Holmes, & Roche, 2001a). That is, an individual may learn that B is "More-than" A, and C is "More-than" B, and in the absence of further training, may derive that C is "More-than" A, and A is "Less-than" C. This is similar to studies examining TI in that following training on a number of adjacent stimulus pairs, participants may respond to the relation between non-adjacent stimuli, in the absence of further training. However, a critical difference between current associative learning, behavioural, and cognitive accounts of TI, and those based on multiple stimulus relations (e.g., Relational Frame Theory; Hayes et al., 2001a), centres on the role of verbal behaviour. For example, Relational Frame Theory (RFT) proposes that a history of training across multiple exemplars of contextually controlled arbitrarily applicable relations (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2001) is central to humans' ability to demonstrate verbal behaviour. Indeed, evidence is gathering showing that verbally sophisticated humans readily demonstrate derived stimulus relations, while humans with language impairments, and non-humans, have not yet demonstrated convincing evidence of these relations (e.g., Barnes, McCullagh, & Keenan, 1990; Dugdale & Lowe, 2000; Dymond, Roche, & Barnes-Holmes, 2003; Hayes, 1989). Such findings have been taken as evidence linking derived stimulus relations to language development (Gross & Fox, 2009). The following section will look at derived stimulus relations in more detail.

Derived Relational Responding

As mentioned, the emergence of novel or untrained relations between non-adjacent stimuli such as that seen in a TI task is directly comparable to behaviour-analytic research on derived relational responding. That is, when verbally-able humans are trained on a number of adjacent conditional discriminations, untrained but predictable relations often emerge between non-adjacent stimuli (e.g., Dymond & Whelan, 2010; Hyland, O'Hora, Smyth, & Leslie, 2012; Reilly, Whelan, & Barnes-Holmes, 2005; Whelan, Barnes-Holmes, & Dymond, 2006; for a review see

Dymond, May, Munnelly, & Hoon, 2010). For example, in a conditional discrimination, participants are presented with two discriminative stimuli (S+ and S-) in the presence of a conditional stimulus (Saunders & Williams, 1998). Thus, the function of the discriminative stimulus changes on the basis of the conditional stimulus presented.

The earliest findings in support of this have emerged from studies of stimulus equivalence (e.g., Sidman, 1971). For example, when examining the emergence of this pattern of behaviour in the laboratory, participants may be presented with one of two sample stimuli, along with two comparison stimuli (Barnes-Holmes & Barnes-Holmes, 2000). Therefore, when attempting to establish responding in accordance with equivalence relations, participants may first learn to choose A in the presence of B (i.e., AB) and B in the presence of C (i.e., BC). It then follows that a number of untrained relations are likely to emerge between B and A, and C and B (symmetry), A and C (transitivity) and C and A (combined symmetry and transitivity; see Figure 1.3). Derived stimulus relations such as those described in the above example are not limited to the phenomena of stimulus equivalence. A growing body of literature provides evidence that humans can learn to respond in accordance with a variety of derived stimulus relations that include: "Same" and "Opposite" (e.g., Dymond, Roche, Forsyth, Whelan, & Rhoden, 2007, 2008; Dymond & Whelan, 2010; Whelan & Barnes-Holmes, 2004); "More-than" and "Less-than" (e.g., Berens & Hayes, 2007; Dymond & Barnes, 1995; Gorham, Barnes-Holmes, Barnes-Holmes, & Berens, 2009; Munnelly, Dymond, & Hinton, 2010; O'Hora, Roche, Barnes-Holmes, & Smeets, 2002; Vitale, Barnes-Holmes, Barnes-Holmes, & Campbell, 2008), and "Before" and "After" (e.g., Hyland et al., 2012; O'Hora, Roche, Barnes-Holmes, & Smeets, 2004).

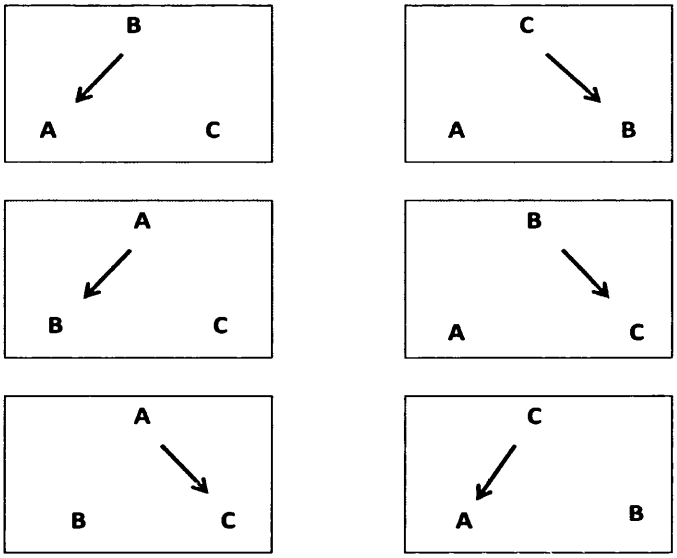


Figure 1.3. An example of a matching-to-sample task commonly employed throughout research on stimulus equivalence. The upper panel displays trials presented during the initial training phase (i.e., B-A and C-B). The middle panel is an example of probes for symmetry (i.e., A-B and B-C), while the lower left panel shows a probe for transitivity (A-C), and the lower right panel shows a probe for equivalence (C-A).

The aforementioned examples of derived stimulus relations represent specific instances of *arbitrarily applicable relational responding* (AARR) that are termed relational frames. Relational frames in turn are defined on the basis of three major properties: *mutual* and *combinatorial entailment* and the *transformation of functions*. The first, *mutual entailment* is the generic term for the concept of symmetry in stimulus equivalence and refers to the derived bidirectionality of a stimulus relation. For example, if A is MORE-THAN B, then a LESS-THAN relation is mutually entailed between B and A (i.e., B is LESS-THAN A). The term *combinatorial entailment* corresponds to "transitivity" and "equivalence" in stimulus equivalence, and refers to the emergence of a derived relation in which two or more stimulus relations mutually combine (e.g., Fields, Adams, Verhave, & Newman; 1990; Hayes et al. 2001a; Sidman, 1971; Torneke, 2010). For example, if A is MORE-THAN B, and B is MORE-THAN C, a "More-than" relation emerges between A and C (i.e., A is MORE-THAN C), and a "Less-than" relation is entailed between C and A (i.e., C is LESS-THAN A). A *transformation of stimulus functions* refers to the fact that when a given stimulus within the relational network acquires a certain psychological function, the functions of other stimuli within the network may also be altered based on the underlying relation between them (Hayes, Fox, Gifford, Wilson, Barnes-Holmes, & Healy, 2001b). By fulfilling the preceding three requirements, all of which are established through operant learning (i.e., patterns of behaviour controlled by consequences; Torneke, 2010), the definition of arbitrarily applicable derived relational responding is fulfilled (Berens & Hayes, 2007). The method in which arbitrary relational responding is considered a generalised operant will become clearer in the following paragraphs. However, it is the study of derived relational responding to the comparative frames of "More-than" and "Less-than" that is most relevant to the investigation of TI.

In order to train and test comparative and indeed other types of relational frames in the laboratory, a number of important experimental training and testing procedures have been developed. Specifically, two separate training and testing phases, namely non-arbitrary and arbitrary relational training and testing have been incorporated throughout numerous studies with both children and adults (e.g., Berens & Hayes, 2007; Gorham et al., 2009; Munnelly et al., 2010; O'Hora et al., 2002; Reilly et al., 2005; Whelan et al., 2006). Non-arbitrary relational training consists of presenting multiple exemplars of differing stimulus sets, where participants learn to respond to "one event in terms of the other based solely on the formal properties of the related events" (Hayes et al., 2001a, p. 25). So, for example, a participant may be presented with two stimuli that differ in terms of a specified physical dimension, such as quantity (e.g., one apple and three apples), and correct selections of three apples would be reinforced in the presence of the MORE-THAN contextual cue, and selections of one apple in the presence of the LESS-THAN contextual cue. Following an appropriate history of explicit reinforcement across a number of differing stimulus sets, this pattern of responding may generalise when an appropriate test for the emergence of non-arbitrary comparative responding is presented, involving novel stimulus sets.

Numerous studies have demonstrated that both humans and non-humans are capable of responding to the non-arbitrary relations between stimuli. However, humans are also capable of responding to the relations between stimuli under the control of contextual features that are not dependent on the physical properties of the relata (Hayes et al., 2001a). That is, humans can learn to respond to stimuli that are physically dissimilar. For example, one initially learns to correctly respond to stimuli that are physically similar in the presence of the contextual cues for MORE-THAN and LESS-THAN, and for humans, this responding may also extend and generalise to arbitrary stimuli encountered in the appropriate relational context (Hayes et al., 2001a, p.25). This is highlighted in the following example in which a child may be presented with a one-euro and 50-cent coin. The child may initially select the 50-cent coin, on the basis that it is "bigger-than" the one-euro coin. However, later the child learns that the one-euro coin is "worth" "More-than" the 50-cent coin irrespective of its physical size, and the child will select the one-euro coin. Thus, *arbitrarily applicable relational responding* is controlled by the context, and not the physical properties of the stimuli (O'Hora et al., 2002).

An empirical demonstration of responding in accordance with the relational frame of "More-than" and "Less-than", was first reported by Dymond and Barnes (1995). In the first phase of this experiment, termed non-arbitrary relational training and testing, the contextual functions of SAME, OPPOSITE, MORE-THAN and LESS-THAN were established for four arbitrary images. For example, during a particular training trial, participants were presented with the contextual cue for MORE-THAN, a 6-star sample, and 3-star and 9-star comparisons. On this particular trial, selections of the 9-star comparison were reinforced. During another training trial however, participants were presented with the same sample and comparison stimuli, but in the presence of the LESS-THAN contextual cue. In this instance, selections of the 3-star comparison were reinforced. Participants were exposed to the above training across several stimulus sets of differing quantities. Non-arbitrary relational testing followed, and involved the presentation of novel stimulus sets in the absence of feedback. In a second phase, three contextual cues (SAME, MORE-THAN and LESS-THAN) were used to train six arbitrary relations among stimuli that were physically dissimilar from one another (i.e., consonant-vowel-consonant). The resulting network was as follows: B1 was the SAME AS A1 and C1; B2 was LESS-THAN A1, and C2 was MORE-THAN A1. The three most important emergent relations were: C2 MORE-THAN B1 (because C2 is MORE-THAN A1, which is the SAME AS B1, C2 is MORE-THAN B1), and B2 is LESS-THAN B1 (because B2 is LESS-THAN A1, which is the SAME AS B1, B2 is LESS-THAN B1).

Following this first seminal study, numerous researchers have both replicated and extended these initial findings. For example, Whelan et al. (2006) demonstrated a transformation and generalisation of consequential functions in accordance with the derived comparative relations of "More-than" and "Less-than" with a 7-member relational network (A-B-C-D-E-F-G). Following non-arbitrary relational training and testing, participants were trained on six conditional discriminations (i.e., A<B, B<C, C<D, E>D, F>E and G>F), and this was followed by a test for all possible derived relations (e.g., B<F). In the next phase, stimulus D was paired with the delivery of points, and a test for the transformation of consequential functions involving stimuli within the relational network followed. The authors found that participants always selected the highest-ranked consequential stimulus in the network (e.g., G) when it was presented alongside the other stimuli from the relational network in a simultaneous discrimination. Thus, these findings replicate and extend Dymond and Barnes' (1995), from a 3- to 7-member relational network, and from the transformation of self-discriminative functions to the transformation of consequential functions.

In a further study, Dougher, Hamilton, Fink and Harrington (2007) demonstrated the transformation of eliciting and discriminative functions to the derived comparative relations of "More-than" and "Less-than". In one experiment, using a conditional discrimination protocol, participants were trained to select the smallest, middle, and largest member from a series of three-comparison arrays, in the presence of three samples, A, B and C. The "middle" stimulus B was then selected and used, to train a steady rate of keyboard pressing. This was then followed by the presentation of stimulus A (smallest) and stimulus C (largest). The authors found that participants emitted keyboard responses slower to A and faster to C, than they emitted to B. Stimulus B was then paired with a mild electric shock, and measures of changes in skin conductance were recorded as the dependent variable. As expected, the authors found that skin conductance changes were smaller for A, and larger for C, relative to those for B. In a subsequent experiment, stimulus A was used as the sample to establish arbitrary size rankings among four coloured circles of the same size. A middle circle was then selected to train a steady rate of keyboard pressing, before the introduction of the remaining circles. The authors found that responses to the "smaller" and "larger" circles were slower and faster, respectively, relative to the "middle" circle. Thus, the results of these experiments demonstrated that the derived relations established among stimuli during training, allowed participants to correctly infer the relative size rankings among novel sets of stimuli.

The foregoing studies all demonstrate that it is possible to establish derived comparative responding and the transformation of stimulus functions to 3- and 7-member relational networks. However, of more relevance to research on TI, is a study by Reilly et al. (2005), in which response latencies to stimuli in a 5-member relational network were examined. This study is particularly suited to an analysis of TI-like behaviour as it permits an analysis of the symbolic distance effect (SDE) between response latencies to baseline, mutually entailed, and one-, two- and three-node combinatorially entailed relations. For instance, as mentioned earlier, the SDE predicts that response times decrease as the number of intervening items between the stimuli in a test pair increases (Moyer & Bayer, 1976). With respect to the Reilly et al. (2005) study, participants were first exposed to non-arbitrary relational training and testing, followed by arbitrary relational training and testing (see Figure 1.4). More specifically, during arbitrary relational training, participants were randomly assigned to one of three training groups: All-More, All-Less or Less-More, in which the arbitrary relations trained differed for each group. That is, for the All-More training group, the baseline relations consisted of "More-than" relations (B>A, C>B, D>C and E>D); for the All-Less training group, the baseline relations consisted of all "Less-than" relations (A<B, B<C, C<D and D<E); and for the Less-More training group, a combination of "More-than" and "Less-than" baseline relations were presented during training (A<B, B<C, D>C and E>D). All three groups were then exposed to the same arbitrary relational test in which novel stimulus combinations were presented. In this test phase, the baseline relations were tested without feedback, along with 16 novel test pairs which included four mutually entailed relations (All-More: A<B, B<C, C<D and D<E; All-Less: B>A, C>B, D>C and E>D, and Less-More: B>A, C>B, C<D and D<E), six one-node (C>A, D>B, E>C, A<C, B<D and C<E), four two-node (D>A, E>B, A<D and B<E), and two

three-node (E>A and A<E) combinatorially entailed relations. A node in this instance, as defined by Fields, Verhave and Fath (1984) is a stimulus that is linked by training at least two other stimuli. Therefore, when probing for emergent relations, such as BD, the C stimulus is the node that separates the two stimuli. The number of nodes that separate the stimuli between which an emergent relation is formed can vary (e.g., one-, two- and three-node relations), and this is referred to as the nodal distance. Thus, in the Reilly et al. (2005) study, participants were trained and tested on a 5-member relational network, and when exposed to probes for one-node relations, there was one intervening stimulus as the node (e.g., D>B, with C as the node). During probes for two-node relations, there were two intervening stimuli as the nodes (e.g., E>B, with C and D as the node), and during probes for three-node relations there were three intervening stimuli as the nodes (e.g., E>A, with B, C and D as the node). Reilly et al. (2005) found that response latencies decreased linearly from the baseline to three-node combinatorially entailed relations. That is, the slowest response latencies were observed on the baseline relations, with the fastest on the three-node combinatorially entailed relations. In addition, the authors also found that response latencies to all types of relations were significantly faster for the All-More training group relative to the other training groups.

An additional study by Munnelly et al. (2010) examined derived comparative performances at test using procedures similar to Reilly et al. (2005). However, in this study, the authors were concerned with performance accuracy rather than response latencies, and performances for only two training groups (All-More and All-Less) were compared. Findings demonstrated that high accuracy was maintained on the baseline relations, and no differences were observed between the groups in terms of accuracy on novel inferential pairs. However, a finding that was of interest, was, that for the one- and two-node combinatorially entailed relations, accuracy was significantly higher on the relations that were the *same as training*, in comparison to those that were *different to training*. A *same relation as trained* in this instance can be seen for the All-More group who were trained on only "More-than" baseline relations (e.g., B>A, C>B, D>C and E>D), but tested on a combination of "More-than" and "Less-than" relations. The one node C>A test trial would be considered the *same as training* as it was a "More-than" relation, whereas the A<C one-node test trial would be considered *different to training as* it was a "Less-than" relation.

This finding was observed for both the All-More and All-Less groups, and thus, the authors concluded that relations that are different to the training, may involve a greater response effort, and so, are more difficult to solve.

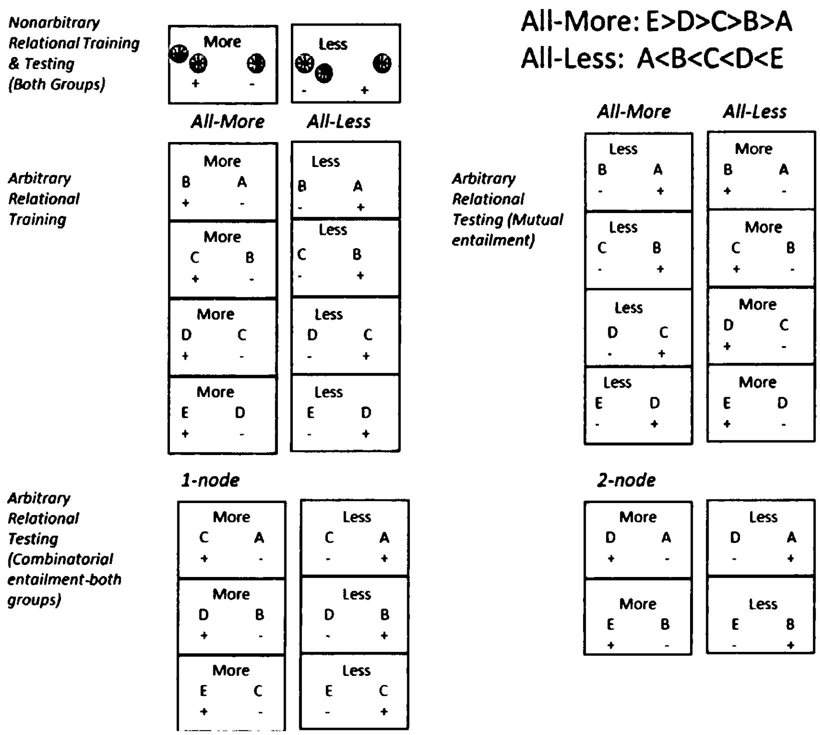


Figure 1.4. An example of the non-arbitrary and arbitrary relational training and testing phases that participants were exposed to in the Reilly et al. (2005) study.

A further study to examine TI behaviour with adult participants was conducted by Vitale et al. (2008). In this study, participants were exposed to a number of "More-than" and "Less-than" problem-solving tasks involving a 3-term series. During the experimental task, participants were exposed to six trial types in which the target relations among the stimuli were either, explicitly stated (e.g., problems that present the *same* comparative term within the premises), specified/unspecified (e.g., problems that have a clear or unclear solution), or transitive (e.g., non-linear problems that require conversion to linear relations before they can be solved). Each trial consisted of three identically sized containers (termed "coins") being presented onscreen, where the coins differed only in terms of colour (e.g., A = red, B = blue and C = yellow). Three brown containers (known as coffee containers) were also presented onscreen. However, the coffee containers differed in terms of size (e.g., large = “full of coffee”, medium = “half full of coffee” and small = “little coffee”). In addition, a small black tin was presented onscreen, known as the “I cannot know” tin. During a particular trial, participants were required to determine the relations between all three coins. For example, during a trial in which the relations were stated among the stimuli, participants were presented with the following instructions onscreen: “The red (A) coin is worth LESS-THAN the blue (B) coin, and the blue (B) coin is worth LESS-THAN the yellow (C) coin (e.g., $A<B$ and $B<C$). A correct response consisted of the participant deriving that the red (A) coin is worth the least and thus, placing it in the smallest coffee container. The blue (B) coin was in the middle, and so should be placed in the medium container, while the yellow (C) coin is the largest, and so should be placed in the largest container. On the other hand, during an unspecified-mixed trial (containing both linear and non-linear trial types), such as $C>B$ and $B<A$, participants were instructed that: “The yellow (C) coin is worth MORE-THAN the blue (B) coin, and the blue (B) coin is worth LESS-THAN the red (A) coin”. In this instance, a correct response consisted of participants deriving that the blue (B) coin is worth the least, and thus, placing it in the smallest coffee container. However, for the remaining coins, it is impossible to derive the relationship between the red (A) and yellow (C) coins, and thus, both coins should be placed in the “I cannot know” tin. Findings revealed that participants were successful in displaying the targeted “More-than” and “Less-than” relations. However, for some of the more difficult problem-solving tasks, accurate responding to these relations only emerged once participants were exposed to interventions that targeted these deficiencies (Vitale et al., 2008).

Training and Testing Protocols

As can be seen from the above studies, the method in which derived comparative relations are examined in the laboratory can vary. For instance, the training and testing protocol employed by Dougher et al. (2007) to establish derived comparative responding to a 3-member relational network, differed from those employed by Vitale et al. (2008). Furthermore, the procedures employed by Reilly et al. (2005) and Munnelly et al. (2010) also differed from the other two studies in that participants were first to exposed to non-arbitrary relational training and testing followed by arbitrary relational training and testing. Despite the differences across studies, each protocol involved participants being exposed to all test relations concurrently. That is, in the Reilly et al. (2005) study for example, participants were exposed to both mutually and combinatorially entailed relations in the same test block, while, similarly, in the Vitale et al. (2008) study, participants were exposed to problem-solving tasks of varying levels of difficulty, at the same time. However, as findings from both the Vitale et al. (2008) and Munnelly et al. (2010) studies demonstrate, some of the targeted relations emerge more readily than others, and thus, it may be beneficial to examine the perquisites necessary for the emergence of this type of responding in an experimental context. For instance, studies examining equivalence relations have noted facilitative effects on the emergence of this type of responding, when a simple-to-complex testing protocol was incorporated into the experimental design (e.g., Adams, Fields, & Verhave, 1993; Fields, Varelas, Reeve, Belanich, Wadhwa, de Rosse, & Rosen, 2000). More specifically, with this protocol, probes for the properties of symmetry are presented before tests for equivalence, and thus, some researchers propose that it is this sequential presentation of the necessary prerequisites, which enhances performance on equivalence relations at test (e.g., Adams et al., 1993; Fields, et al., 2000).

However, as seen thus far, no studies examining the emergence of derived comparative responding with adult participants have sought to explicitly examine the prerequisites that are necessary for the emergence of this behaviour. Thus, the question could be posed as to whether successful responding in accordance with the properties of mutual entailment is central to, and facilitates the emergence of combinatorially entailed responding. This in turn seems to be an important issue, as, when the predicted relational performances fail to emerge, it may be necessary to develop interventions aimed at targeting these weaknesses.

Interventions

Another point worth considering when examining the emergence of derived comparative responding in the laboratory is the availability of interventions when the predicted relational repertoires are weak, or do not emerge immediately. For example, when Vitale et al. (2008) found that adult participants in their study were displaying weak responding on some of the targeted relations, the authors examined the effectiveness of automated feedback, repeated test exposures, and the use of non-arbitrary trials, in generating this behaviour. Vitale et al. first examined the effectiveness of repeated test exposure by exposing participants to the test protocol for a total of six times. Although findings demonstrated that relational performances were strong on the specified relations, weaknesses were observed on some of the unspecified and unspecified transitive relations, and thus, repeated test exposures did not lead to the desired improvements in performances. The authors next examined the utility of written feedback, where the word "Correct" appeared onscreen if the participant responded correctly, and the word 'Wrong" appeared if they responded incorrectly. Results demonstrated that performances were again strong on the specified relations and improvements were also noted on the unspecified relations. However, performances still remained weak on the unspecified transitive relations, and therefore, Vitale et al. explored the efficacy of a non-arbitrary training intervention. With this intervention, the targeted deficient relations were highlighted in their non-arbitrary form. That is, the actual size of the three coins was altered so that they were no longer identical in size, and thus, the targeted relations were no longer arbitrary. Results demonstrated that performances on the more difficult unspecified transitive relations were still weak, even when participants were exposed to an increased number of non-arbitrary training trials. Finally, Vitale et al. explored whether a combination of automated feedback and non-arbitrary training trials would lead to the desired improvements on the weaker targeted relations. Although findings demonstrated that stronger improvements in performances were noted when both types of interventions were combined, some weaknesses were still observed on the unspecified relations.

The findings from the Vitale et al. (2008) study highlight the importance of the development of interventions to tackle weaknesses in relational repertoires. However, prior to this study, relatively few studies have examined the effectiveness of interventions with adult populations, with the majority of research stemming from studies involving young children (e.g., Y. Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004; Berens & Hayes, 2007; Gorham et al., 2009). An example of one such study is that by Y. Barnes-Holmes et al. (2004), in which multiple-exemplar training (MET) was employed as an intervention to establish arbitrarily applicable comparative responding in young children. In this study, three children were exposed to a number of "More-than" and "Less-than" problem-solving tasks, involving two or three coins (A-B-C). During initial baseline tests, participants were informed about the value of each coin, and then asked which they would use to buy candy. Results demonstrated that during baseline tests, accuracy was below chance (i.e., 50%) on both mutually and combinatorially entailed relations. Thus, Barnes-Holmes et al. (2004) exposed the children to a multiple-exemplar training intervention, which consisted of training and testing across multiple examples of "More-than" and "Less-than" relations. Findings demonstrated a considerable improvement in participants' ability to respond to the properties of mutual and combinatorial entailment following the MET intervention.

Berens and Hayes (2007) conducted a similar study, which aimed to replicate and extend these findings. However, in this study, the authors were concerned with the sequence, and amount of training that leads to the emergence of arbitrary comparative responding in young children who were deficient in this relational repertoire. Throughout the study, children were exposed to both linear and non-linear trials. For example, during a trial in which the stimuli were presented in a linear order (e.g., A>B), participants were told that "This [pointing to Picture A] is MORE-THAN that [pointing to Picture B]". On the other hand, during a mixed non-linear trial (e.g., A>B>C), participants were told that "This [pointing to A] is worth MORE-THAN that [pointing to B], and this [pointing to C] is LESS-THAN that [pointing to B]". The following question was then presented to children after each trial: "Which would you use to buy candy?" Results demonstrated that reinforcement across multiple exemplars of "More-than" and "Less-than" relations, facilitated the emergence of arbitrary comparative responding, and that these skills generalised across both stimulus sets and trial types. In addition, this study also identified that responding in accordance with non-arbitrary comparative relations is a necessary prerequisite to the emergence of arbitrary comparative responding for young children.

The studies outlined above share a general consensus that an appropriate history of responding in accordance with a range of non-arbitrary comparative relations is an essential precursor to the ability to display arbitrary comparative relational responding. In addition, these studies also appear to provide further support for the view, that a history of MET across numerous stimulus sets facilitates the emergence of this behaviour (see also Gorham et al., 2009). However, one issue that remains is whether the interventions employed in studies involving young children will generalise to adult populations, and to individuals with developmental delays. In addition, few studies currently among the literature on TI have focused on interventions or methods for improving performances on training and inferential pairs, when these are found to be weak or deficient. For example, a study by Smith and Squire (2005) compared performances for memory-impaired patients and a group of controls on the TI task. Findings revealed that the memory-impaired individuals performed poorly on the transitive pair BD. Even more interesting was the finding that these patients were unable to reach criterion on the premise pairs following two consecutive days of training. A further study by Ryan et al. (2009) compared performances for younger and older adults on a number of inferential test pairs. Findings revealed that weaknesses were noted on learning of the premise pairs, accuracy on inferential test, and response times for participants that were suffering from age-related deficits. Thus, an important issue resulting from both of these studies is the lack of appropriate interventions that may be incorporated in an attempt to remediate these weaknesses or deficits. In contrast, a core objective of RFT, and indeed behaviour analysis, is to examine the success of a range of behavioural interventions in establishing or facilitating the emergence of relational repertoires (Vitale et al., 2008), which in turn may be beneficial for populations that display weak relational repertoires.

Features of the Current Account of TI as Derived Comparative Relational Responding

The purpose of the current thesis is to explore a variety of behavioural protocols and paradigms based on Relational frame theory, to establish an alternative account of TI-like behaviour with adult participants. (Note. The definition of "TI-like" behaviour has been adopted throughout the current thesis as responding to a combination of "More-than" and "Less-than" relational problems is examined. Currently, studies examining TI can only present either "More-than" or "Less-than" test problems, and thus, the term "TI-like" is used to describe the acquisition of "More-than" and "Less-than" derived comparative relations). One of the features of the current account is that it permits a comparison of the effects that differing training structures may have on subsequent relational responses at test. For example, participants are assigned to one of two, or three training groups (e.g., All-More, All-Less and Less-More), and it is possible to compare participants' responses to both "More-than" and "Less-than" relations at test, following exposure to arbitrary relational training involving only all "More-than" (the All-More group), all "Less-than" (the All-Less group), or a combination of "More-than" and "Less-than" (Less-More) relations. Previous findings have suggested that "More-than" relations develop earlier than "Less-than" in the repertoires of young children (e.g., Barnes-Holmes et al., 2004), and that reaction times are faster at test for participants assigned to an All-More training group relative to two other training groups (e.g., All-less and Less-more; Reilly et al., 2005). However, the suggestion that "More-than" relations emerge earlier than "Less-than" relations in a child's repertoire requires further empirical investigation, because as the authors admit, this issue, for the moment is just speculative.

Another distinct feature associated with the current training and testing paradigm is that, in comparison to those currently employed in studies of TI, precise predictions about relational responses at test can be made on the basis of the contextual cue presented. For example, when presented with the test pair CE, in the presence of the MORE-THAN contextual cue, correct selections of E are predicted over C (E>C), and in the presence of the LESS-THAN cue, correct selections of C are predicted over E (C<E). The different training designs also allow for a greater number of inferential trials to be presented at test. For example, for each training group, each trained relation also entails a bidirectional relation, which can be presented under the appropriate contextual cue at test. So, the All-More group, who are trained on B>A, C>B, D>C and E>D, may be tested on the following mutually entailed relations in the presence of the LESS-THAN cue: A<B, B<C, C<D and D<E. Similarly, for the All-Less group, the trained relations are: A<B, B<C, C<D and D<E, with tests for mutual entailment presented under the contextual cue for MORE-THAN (B>A, C>B, D>C and E>D). However, with the training and testing designs that are currently employed in studies of TI, it is not possible to train participants on, for example, B>A and then test for A<B. Thus, the contextual

control exerted over training and testing allows for greater control and more precise predictions to be made about relational responses.

Also, when probing for one- and two-node combinatorially entailed relations, it is again possible to test for both "More-than" and "Less-than" relations, regardless of group assignment. Thus, both groups may be presented with the same one- (C>A, D>B, E>C, A<C, B<D and C<E) and two-node (D>A, E>B, A<D and B<E) trials at test. The combinatorially entailed relations may also be further broken down into same *relation as trained* and *different relation as trained,* on the basis of a distinction drawn by Reilly et al. (2005). To illustrate, participants in the All-More group, for example, may be presented with one-node relations that are the *same as training* (e.g., C>A, D>B and E>C), but also with test trials that are *different to training* (e.g., A<C, B<D and C<E). This distinction also applies to the two-node relations.

The current protocol has the ability to examine the emergence of derived comparative responding by employing an experimentally manipulated history of reinforcement (Berens & Hayes, 2007; see also Munnelly et al., 2010). This in turn provides a direct test of the emergence of this pattern of responding in the laboratory setting.

RFT versus behavioural and cognitive accounts of TI

The behavioural account proposed in the current thesis offers an alternative to both basic behavioural and associative learning, and higher-order cognitive accounts of TI. For example, the current approach proposes that a history of multiple-exemplar training across different types of "More-than" and "Less-than" relations in many different contexts is central to the emergence of derived comparative responding. Associative learning, and behavioural theories of TI also propose that the learning context experienced during training is important in the emergence of TI. However, the strategies employed to solve transitive problems differs between associative learning and RFT accounts of TI. For example, VTT proposes that values accrued during training, account for correct selections at test (von Fersen et al., 1991). That is, during training, value transfers from the reinforced stimulus to the non-reinforced stimulus, and thus, at the end of training, each stimulus will end up with a different value (e.g., A = +2, B= +1, C= 0, D = -1, E = -2; see Libben & Titone, 2008). According to VTT, correct selections of, for example, B over D at test, are made on the basis of the differential values both stimuli accrue during training (e.g., B = +1, D = -1). In contrast, from an RFT perspective, contextual cues change the function of the discriminative stimulus, and thus, account for selections at test. For example, when stimuli C and A (C>A) are presented in the presence of the MORE-THAN contextual cue, correct selections of stimulus C may be predicted, whereas, when both stimuli are presented in the presence of the LESS-THAN contextual cue, correct selections of stimulus A may be predicted (A<C). In addition, RFT proposes that the ability to respond to derived comparative relations is dependent on verbal behaviour. For instance, there is growing evidence linking the ability to respond to arbitrary relations, such as the frame of comparison, to language ability (see Gross & Fox, 2009). Thus, from an RFT perspective, the ability to respond to "More-than" and "Less-than" problem-solving tasks, as seen in the TI problem, is dependent on language. The contextual approach adopted by RFT also differs from that proposed by cognitive accounts of TI. For instance, cognitive theorists propose that, during training, stimuli become integrated into a metaphorical mental representation, and correct selections at test, merely require the individual to abstract the correct stimulus from its appropriate spatial position on this line (e.g., de Soto et al., 1965; Huttenlocher, 1968). In contrast to RFT accounts of TI, the Image model theory does not propose a role for language in the ability to solve task problems. Thus, all three approaches differ with respect to the strategies employed to solve inferential problems are solved at test.

Furthermore, all three theories differ in terms of predictions at test. For example, VTT proposes that, in a 5-term series (A-B-C-D-E-F), the baseline pairs AB and DE, which contain end terms (e.g., A and E), will be solved more accurately than the baseline pairs that do not contain end terms (e.g., BC and CD). VTT also proposes that any test pair, containing an end term (e.g. AC, AD, AE), will be solved more accurately than those that do not contain end terms (e.g., BD; see von Fersen et al., 1991). For example, von Fersen et al. (1991) propose that as stimuli A and E possess either the strongest or weakest associative strengths, accuracy on test trials containing either of these stimuli, will be superior in comparison to those that do not contain end terms (e.g., BD). In addition, VTT proposes that, superior performances will be observed on the inferential non-endpoint test pair, BE over the BD test pair, in a 6-term series. For instance, Frank et al. (2003, 2005) propose that during training, stimulus B acquires a net positive value (e.g., +1), while stimulus E acquires a net negative value (e.g., -2). The authors propose that the different values accrued on these stimuli, is sufficient to account for correct selections of B over E at test. Furthermore, Frank et al. (2003, 2006) propose that the smaller difference in values between stimulus B (e.g., +1) and D (e.g., -1), account for the lower levels of accuracy observed on the BD test pair. The Image theory also proposes that higher accuracy will be observed on the BE test pair over the BD test pair. For instance, the symbolic distance effect (SDE; Moyer & Bayer, 1976) proposes that accuracy increases, and response times decrease, when there are more intervening items, between the stimuli in a test pair. That is, test pairs with a symbolic distance of one (e.g., BD), require greater manipulation of the mental line, than those with a symbolic distance of two (e.g., BE), and thus, on the basis of these proposals, more accurate performances on the BE test pair, are predicted (e.g., Acuna et al., 2002; Bond et al., 2003; Frank et al., 2003, 2005).

From an RFT perspective, and similar to proposals by VTT, researchers proposes that test pairs containing the end items A and E, will produce higher accuracy, and faster reaction times, than those that do not (see Reilly et al., 2005). For example, during training, stimulus A functions as a stronger S+ for reinforcement than the other stimuli in the network (B, C, D and E), as it is always correct. Similarly, stimulus E is the strongest S- during training, as it is always incorrect. Thus, test trials containing these end terms (e.g., C>A, D>A, E>B), will produce faster reaction times and higher accuracy than those that do not (e.g., D>B and B<D). In addition, RFT proposes that, on this basis, reaction times will differ between one-, two-, and three-node relations. For instance, according to Reilly et al. (2005), one-node relations contain two test pairs that do not contain an end term (e.g., D>B and B<D), while all the test pairs in the two- (D>A, E>B, A<D and B<E) and three-node (E>A and A<E) relations contain an end term. Thus, similar to the SDE, Reilly et al. (2005) propose that the longest reaction times should be observed on the one-node relations, with the shortest on the three-node relations.

The protocols employed throughout the current thesis to establish derived comparative responding as an account of TI in adult humans also differ from both associative learning and cognitive approaches. For example, both associative learning and cognitive accounts employ a simultaneous discrimination to examine the TI problem, whereas RFT employs a conditional discrimination. The difference between both protocols centres on the method of stimulus presentation. For example, with the simultaneous discrimination, two or more discriminative stimuli are presented simultaneously, where one stimulus is the reinforced response (S+) and the other is the non-reinforced response (S-). With the conditional discrimination, again, two or more discriminative stimuli are presented, but in the presence of a conditional stimulus (Saunders & Williams, 1998). In this instance, the functions of the discriminative stimulus changes on the basis of the conditional stimulus presented. For instance, if stimuli A and C (A+C-) are presented with the LESS-THAN cue, selections of A may be predicted, and in the presence of the MORE-THAN cue, selections of C may be predicted (C+A-). Thus, the discriminative stimulus changes on the basis of the contextual cue presented. Furthermore, the simultaneous discrimination does not allow associative learning or cognitive accounts to examine "More-than" and "Less-than" relations simultaneously. That is, learning A+/B- does not logically entail B+/A-, and thus, by employing the RFT protocol, we have the potential to examine and generate a larger relational repertoire consisting of a combination of "More-than" and "Less-than" relations simultaneously.

A further difference between RFT and cognitive accounts of TI is that, from a behavioural and RFT perspective, the development and implementation of interventions that may be employed to strengthen relational repertoires is a primary research aim. For example, if an individual initially displays weak responding on some test problems, RFT researchers seek to determine failures in stimulus control that may have contributed to these findings. That is, stimulus control over responding may have been lacking, and the implementation of certain training interventions (e.g., non-arbitrary relational training and repeated exposure to training and test phases) seek to strengthen this control. Indeed, previous research has reported that the targeted comparative relations emerged following the implementation of training interventions (e.g., Gorham et al., 2009; Vitale et al., 2008). For example, Gorham et al. (2009) exposed children with and without a diagnosis of autism to a training intervention consisting of feedback (e.g., "Correct" or "Wrong") on the targeted relations. The authors found that, following the intervention, all children demonstrated the targeted transitive relations (e.g., BD). In addition, Vitale et al. (2008) found that feedback along with a non-arbitrary training intervention was effective in generating the targeted "More-than" and "Less-than" relations in adult participants. In contrast, cognitive accounts of TI do not focus on the development or implementation of training interventions if transitive responding is weak.

TI and the Real World: Some Limitations

The TI task described in the current thesis is important in the sense that it provides researchers with a method of examining human problem-solving abilities in the laboratory setting. The acquisition of an effective and efficient problem-solving repertoire is important for individuals as they are often faced with problem-solving tasks and decision-making processes on a day-to-day basis. However, there are certain instances in which our problem-solving skills or decision-making processes may not be at optimal levels. For example, when deciding to purchase a new car, an individual may review many of the different car manufacturers and have several discussions with friends and relatives about the advantages and disadvantages associated with each type of car. Nevertheless, when purchasing his or her new car and the salesperson offers a better bargain, the individual turns it down. This phenomenon is known as bounded rationality and is widely studied in economics and politics (e.g., Simon, 1996a, 1996b). For example, Simon (1957) defines bounded rationality as "The property of an agent that behaves in a manner that is nearly optimal with respect to its goals as its resources will allow". In other words, individuals are often required to make decisions under constraints of limited knowledge and time, and thus, are not in a position to calculate the optimal solution, even if this solution is available (e.g., Gigerenzer, 1997). The phenomenon may also be applied to performances on the TI task. For example, individuals may have received sufficient exposure to the premise pairs during training, but when it comes to testing, response accuracy may not always be perfect. This is particularly relevant to responses to the critical BD inferential pair, in that, participants often perform at, or just above chance levels. Thus, applying the concept of bounded rationality to responses on the TI task may help researchers to understand additional variables such as the structure of the environment and time constraints, which may limit individuals' ability to make inferential choices at test. Furthermore, as seen in experimental, laboratory-based studies of TI, if the task environment can be specified to predict rational responses from participants, then it becomes possible for researchers to compare observed behaviours with that expected from rational predictions.

Summary of Experimental Chapters

Chapter 1 provides a review of the literature on transitive inference (TI) and highlights some of the characteristic features associated with this type of behaviour. Chapter 1 also reviews some of the variables that have previously been examined in studies on TI, along with a consideration of the different training and testing protocols employed thus far. In the second half of the literature review, an overview of the proposed behavioural and cognitive theories and accounts of TI is undertaken, followed by a review of the RFT approach adopted in the current thesis.

Chapter 2 sought to examine the more traditional way in which the TI problem has been examined. Thus, all experiments in Chapter 2 employed a simultaneous discrimination paradigm. However, a crucial difference between the method in which TI is currently examined and the method in which it was examined in Chapter 2, is that the current thesis sought to determine the utility of repeated exposure to training and test phases and test mastery criterion, in generating more accurate performances at test. A further aim of Chapter 2 was to examine the effects of pre-experimental instructions on the emergence of TI. Thus, in Experiments 2 and 3, one group of participants received additional instructions at the start of the experiment, while a second group did not (see Greene et al., 2001). The effects of the pre-experimental instructions on response accuracy at test, was then compared between the groups.

Chapter 3 sought to further explore the emergence of TI in adult participant. However, in this chapter, a novel approach based on the principles of Relational frame theory, was employed. More specifically, Chapter 3 sought to determine whether responding to a combination of "More-than" and "Less-than" contextually controlled arbitrary applicable comparative relations, could be employed as a novel account of TI. However, the method in which pairs of stimuli were presented during training and test phases differed from those employed in Chapter 2. For example, throughout all experiments in Chapter 3, participants were exposed to a number of conditional discriminations, where the function of the discriminative stimulus changed on the basis of the contextual cue presented. A further aim of this chapter was to examine the utility of a number of interventions (e.g., repeated exposure to training and test phases, non-arbitrary relational training, observing-response) that may be incorporated to generate more accurate responding.

Chapter 4 again sought to examine the emergence of TI in adult participants, and to determine some of the potential factors that influence the emergence of this behaviour (e.g., linearity of training pairs). In addition, Chapter 4 sought to determine the utility of a novel procedure based on a variant of the Relational Completion Procedure (RCP) in establishing *arbitrarily applicable comparative responding* in adult participants. Thus, throughout all experiments in Chapter 4, the variables thought to be most conducive to the emergence of TI, along with a novel procedure, the RCP, were examined.

Chapter 5 sought to examine the transformation of discriminative functions to a 5-member arbitrary comparative relational network. For example, previous research has shown that, when a function is trained to the "middle" stimulus in an arbitrary comparative relational network, participants may respond "More" to members that are ranked higher in the network, and "Less" to members that are lower in the network, in the absence of feedback (see Dougher et al., 2007). Chapter 5 therefore sought to extend this research from a 3- to a 5-member relational network, and determine the conditions necessary to facilitate this pattern of responding.

Chapter 6 summarises the main findings of each experimental chapter in the current thesis. In addition, Chapter 6 considers some of the strengths and weaknesses of the current findings, and discusses the implications of these findings. Chapter 6 also considers some of the differing theoretical positions described in Chapter 1, along with some suggestions for future research.

Future uses of the current protocols

The protocols employed throughout the current thesis have the potential to examine the emergence of derived comparative responding, or TI-like behaviour, in non-human, or human populations that lack sophisticated verbal repertoires. For instance, if with the current protocols, all verbal instructions were omitted and non-humans failed to respond in accordance with the properties of mutual and combinatorial entailment, then this would further strengthen the proposal that sophisticated verbal behaviour is critical to the emergence of derived relational responding. In turn, the current protocols could potentially be employed as an assessment tool for individuals suffering from language-impairments. That is, if individuals with limited verbal behaviour also failed to respond in accordance with the properties of mutual entailment, then this may allow us to identify the pre-requisites necessary to display arbitrary comparative responding. The current protocol may therefore be beneficial to researchers seeking to develop appropriate interventions to target these deficits.

Furthermore, Chapter 5 sought to examine the transformation of discriminative functions to a 5-member arbitrary comparative relational network, as a potential alternative account of the development and maintenance of clinically significant behaviours, such as anxiety. For example, Chapter 5 examined how individuals may respond to "More" to stimuli ranked higher in the network, and "Less" to stimuli ranked lower in the network. Thus, the emergence of such patterns of responding in the absence of a history of reinforcement may have the potential to account for how individuals come to display increased or decreased levels of fear, or anxiety.

Chapter 2

The Role of Awareness and Repeated Exposures to Training and Testing on the Emergence of Transitive Inference

The TI problem has traditionally been considered a hallmark of logical deductive reasoning (Piaget, 1928; Vasconcelos, 2008). In order to examine the emergence of TI, researchers have to date employed the simultaneous discrimination paradigm. For instance, in a 6-term series, participants may be trained on five adjacent stimulus pairs (e.g., A+B-, B+C-, C+D-, D+E-, E+F-; where "+" and "-" represent the reinforced and non-reinforced selections, respectively), and presented with novel, non-adjacent stimulus pairs (e.g., BD and AE), in the absence of feedback at test. Indeed, this protocol has been successfully employed to examine TI in non-humans, young children and adults (e.g., Acuna et al., 2002; Bond, Wei, & Kamil, 2010; Bryant & Trabasso, 1971; Davis, 1992; Frank et al., 2005; Gazes, Chee, & Hampton, 2012; Gillan, 1981; Greene et al., 2001; Merritt & Terrace, 2011; McGonigle & Chalmers, 1984; Moses et al., 2006; Weaver et al., 1997; Wynne et al., 1992).

As mentioned, the simultaneous discrimination is the typical training and testing protocol employed in studies examining TI in adult participants. With this protocol, two or more stimuli are presented simultaneously, where only one stimulus is reinforced (S+), and the others are not (S-). For example, Frank et al. (2005) employed the simultaneous discrimination to examine TI in adult participants to a 6-term series. Participants were first trained on five adjacent stimulus pairs, followed by a test phase involving the four transitive test pairs, BD, BE, CE, and AF. Similarly, Moses et al. (2006) employed this protocol to examine TI in adult participants, but participants were presented with a greater number of probe trials at test (BD, BE, CE, AC, AD, AE, AF, BF, CF, DF).

Despite variations in studies with respect to the number of transitive pairs presented, at test, there are several methodological features common to human TI studies that may impact on test performance. For instance, in studies examining TI, all inferential probe trials are typically presented in one single test block (e.g., Frank et al., 2005; Greene et al., 2001; Lazareva & Wasserman, 2010). A potential limitation of this approach is that if accurate TI fails to emerge, it is not possible to examine whether additional exposure to training and test phases may facilitate subsequent inferential test performance. Doing so would likely result in improved test accuracy and allow for an examination of the time course of the emergence of inferential abilities.

A further methodological issue involved in TI studies, surrounds the use of pre-determined, accuracy mastery criteria. For example, mastery criterion is rarely employed for the trained or inferential pairs at test (e.g., Acuna et al., 2002; Bond et al., 2010; Dusek & Eichenbaum, 1997; Ellenbogen et al., 2007; Frank et al., 2005; Gillan, 1981; Greene et al., 2001; Lazareva et al., 2001; Lazareva & Wasserman, 2010; Merritt & Terrace, 2011; Moses et al., 2006, 2008; Paxton et al., 2012; Wynne, 1995, 1997). Although studies propose that inferential responding is said to have emerged if an average percentage performance accuracy of between 70% and 100% has been achieved on the transitive pairs (BD, BE, CE; Frank et al., 2005; Greene et al., 2001; Moses et al., 2006), little effort has been made by researchers to incorporate pre-determined mastery criterion during test phases. However, doing so would potentially provide us with more reliable methods of determining the emergence of TI in humans.

Thirdly, awareness is another important methodological factor that may influence inferential test performances. For example, "Awareness" is defined as a conscious understanding that the stimuli can be ordered along a hierarchy, which may be used to make inferential judgements (see Greene et al., 2001; Libben & Titone, 2008; Martin & Alsop, 2004; Moses et al., 2006; Smith & Squire, 2005). Considerable debate exists regarding which strategies individuals employ to solve TI tasks, and whether awareness is both necessary and sufficient for accurate TI to emerge. For instance, and as outlined in Chapter 1, some researchers propose that humans are capable of solving the TI task without recourse to awareness (e.g., Frank et al., 2005; von Fersen et al., 1991), while others argue that conscious awareness of the stimulus hierarchy is necessary (e.g., Lazareva & Wasserman, 2010; Libben & Titone, 2008; Martin & Alsop, 2004; Moses et al., 2006, 2008; Smith & Squire, 2005).

Current Experiments

The current chapter sought to address some of the methodological issues noted in the previous paragraphs. For example, Experiment 1 incorporated a simultaneous discrimination, similar to Frank et al. (2005) and Moses et al. (2006) to examine TI responding in adult participants to a 6-term series, with the following adjustments. For example, Experiment 1 sought to determine whether the incorporation of test mastery criterion and repeated exposure to training and test phases would facilitate inferential responding. Participants were initially exposed to training on five adjacent stimulus pairs (A+B-, B+C-, C+D-, D+E- and E+F-), followed by a test phase involving novel, non-adjacent stimulus combinations, consisting of five baseline pairs (AB, BC, CD, DE and EF), three non-endpoint (inferential) pairs (BD, BE and CE), and seven endpoint pairs (AC, AD, AE, AF, BF, CF and DF; see Table 2.1). A pre-determined test mastery criterion of an average performance accuracy of 80% correct was employed across all test pairs, along with repeated exposure to training and test phases for a pre-determined number of times (e.g., 4).

Experiments 2 and 3 were similar, with the exception that the role of awareness on the emergence of TI was also explored. That is, performance accuracy on transitive pairs was compared between participants that were informed that the stimuli could be arranged in a hierarchy (Informed), against a group of participants that were not given this information (Uninformed).

Experiment 1

Method

Participants

Twenty-nine students, seven male and twenty-two female, ranging in age from 19 to 22 years (M_{age} = 19.9, *SD* = .82) were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the study. Ethical approval was obtained from the Psychology Department Ethics Committee before research commenced. Each participant was provided with an information sheet (see Appendix 1) and consent form (see Appendix 2) at the beginning of the experimental task outlining what the task would entail and their rights as a participant. A debriefing form was also distributed on completion of the study (see Appendix 3).

Apparatus and Stimuli

The stimuli employed in the study were six images randomly selected from the Kanji script (see Figure 2.1). The experimental task was programmed in E-prime (version 1.2), which controlled the presentation of all stimuli and recorded all responses.

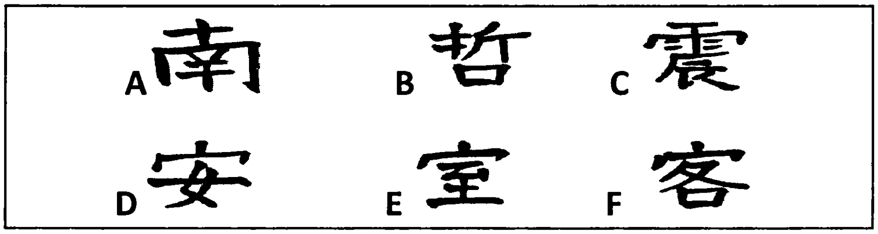


Figure 2.1. The Kanji images employed during the transitive inference training, and test phases. The images are labelled A, B, C, D, E and F (Note: participants were never exposed to these labels).

The experimental procedure consisted of *Phase 1: Transiitve Inference Training* and *Phase 2: Testing.* Participants were first exposed to a transitive inference training phase in which they were trained on five overlapping stimulus pairs (A+B-, B+C-, C+D-, D+E- and E+F-; where "+" and "-" represent the reinforced and non-reinforced selections, respectively; see Table 2.1 and Figure 2.2). Participants were not informed of any relationship between the stimuli and were initially required to learn the pairs by trial and error. On each trial, a pair of images appeared simultaneously, in the middle of the computer screen. To select the image on the right, participants pressed the "m" key on the computer keyboard, while participants pressed the "z" key to select the image on the left. Both images remained onscreen until the participant emitted a response. Left- and right- screen position of the images in each pair was counterbalanced across trials. During training, trials were followed by feedback presented in white on a black background. A correct response was followed by the word "Correct!" displayed in the middle of the computer screen, while an incorrect response was followed by the word "Wrong". Feedback remained onscreen for 1.5 s and was followed by an inter-trial interval (ITI) of 1.5 s.

Table 2.1
Test trials employed during the test phase in Experiments 1, 2 and 3.

	Test trial type						
Baseline pairs	AB	BC	CD	DE	EF		
Endpoint pairs	AC	AD	AE	AF	BF	CF	DF
Non-endpoint pairs	BD	BE	CE				

Procedure

Phase 1: Transitive Inference Training. This phase began with the following instructions onscreen:

> During this phase you will be presented with two images in the middle right- and left-hand side of the computer screen. Your task is to learn to select the correct image. To select the image on the left, press the marked key on the left of the keyboard. To select the image on the right, press the marked key on the right of the keyboard. Sometimes the computer will give you feedback, and at other times it will not. The computer will tell you when this phase of the experiment is finished. Please press the spacebar to begin!

During training, participants were trained on five adjacent stimulus pairs (A+B-, B+C-, C+D-, D+E- and E+F-; where "+" and "-" represent the reinforced and non-reinforced responses, respectively). As mentioned, both of the images from a pair were presented simultaneously in the centre of the computer screen. So, when for example, the training pair AB was presented, selections of A were reinforced with the word "Correct!" while selections of B were unreinforced with the word "Wrong" (A+B-). Similarly, when the training pair BC was presented, correct selections of B were reinforced, while incorrect selections of C were unreinforced (B+C-). Training proceeded in this same manner for the remaining three pairs CD (C+D-), DE (D+E-) and EF (E+F-; see Figure 2.2). All training trials were then followed by an ITI of 1.5 s. The five baseline pairs were presented in a quasi-random order, four times each, within a block of twenty training trials. In order to successfully complete the arbitrary relational training phase, participants were required to achieve 100% accuracy (i.e., make 20 out of 20 correct responses) on a given training block. Training blocks were repeated, if necessary, until this criterion was met.

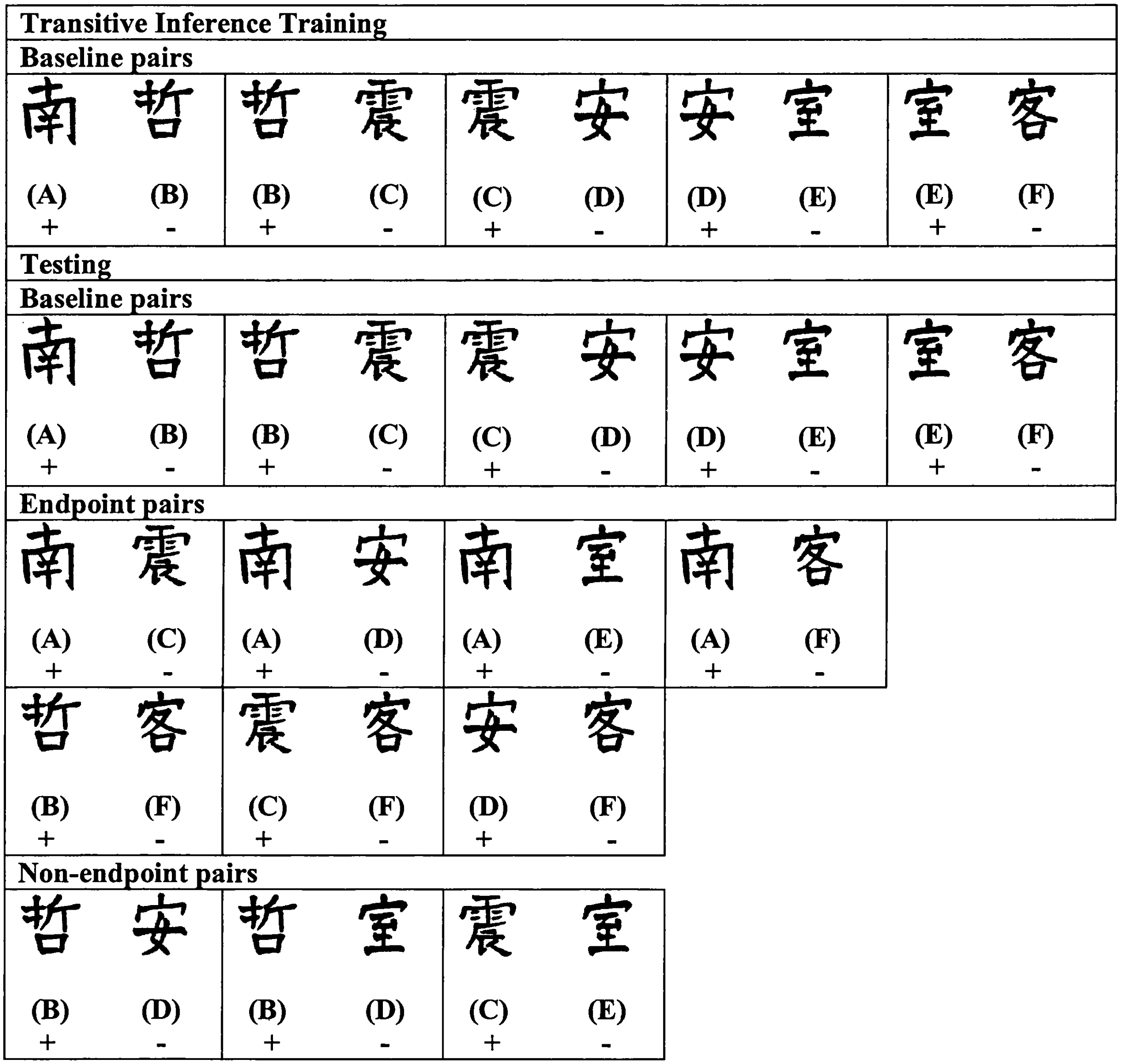


Figure 2.2. An overview of the training and test pairs presented in Experiments 1-3, where "+" and "-" represent the reinforced and non-reinforced responses, respectively.

Phase 2: Testing. On reaching training criterion, participants proceeded immediately to the test phase, where all feedback (i.e., "Correct!" and "Wrong") was now omitted. During this test phase, participants were presented with probes for the maintenance of the five baseline pairs, alongside probes for seven endpoint and three non-endpoint pairs (see Figure 2.2 and Table 2.1). Each test pair was presented four times, resulting in a total of sixty test trials. In order to meet criterion on the test, participants were required to achieve a minimum mean of 80% accuracy (i.e., 48 out of 60 correct responses), across all test pairs (baseline, endpoint and non-endpoint). If participants failed to meet this criterion, they were re-exposed to transitive inference training, followed again by testing for a maximum of three further times.

Results and Discussion

Of the twenty-nine participants that took part in Experiment 1, seven ended their participation, as they were unable to complete transitive inference training after one hour. One participant ended their participation in the study after two unsuccessful attempts at the test phase. A further eight participants failed to meet criterion following four exposures to testing. However, a total of thirteen participants managed to successfully complete the experimental task with the total number of test exposures required, ranging between 1 and 4 (M = 2.00, SD = 1.28). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Training trials to criterion

In order to meet criterion during transitive inference training, participants required between 20 and 660 trials (M = 156.67, SD = 191.08) to do so (see Table 2.2 for a summary of performance accuracy across each exposure to training and testing, for participants that passed and failed the experimental task).

Accuracy: Testing

Baseline Pairs. A graphical representation of the mean percent correct on the baseline pairs for participants that passed and failed the experimental task can be seen in Figure 2.3. This figure shows that, for participants that passed and failed, accuracy was high on all the baseline pairs. Average accuracy on the baseline pairs ranged between 90% and 100%.

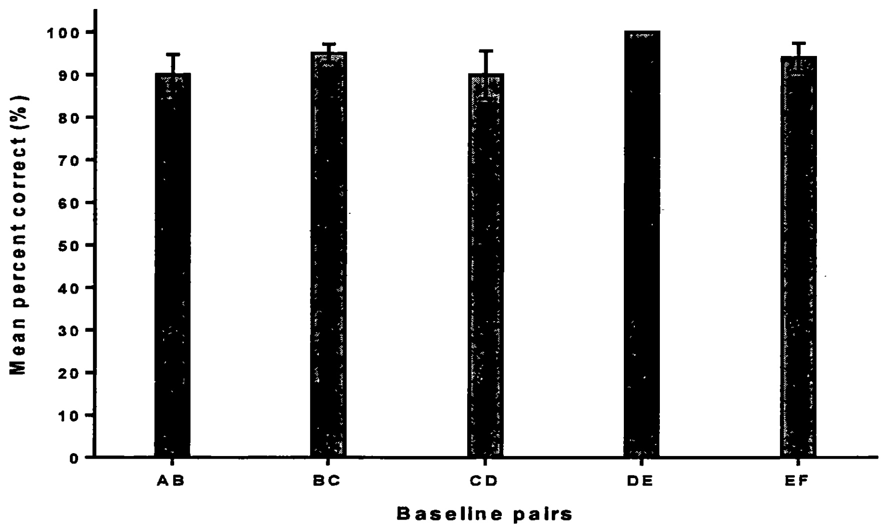


Figure 2.3. Experiment 1. Mean percent correct to the baseline pairs for participants that passed and failed the experimental task. Error bars indicate standard errors.

A summary of response accuracy for participants across each exposure to testing can be seen in Table 2.2. From this table, it can be seen that for the thirteen participants that passed, six (P1, P10, P13, P14, P19 and P28) achieved 100% accuracy on the baseline pairs. For the other seven participants that passed (P8, P9, P11, P15, P17, P20 and P29), accuracy on the baseline pairs ranged between 70% and 90%. Eight participants ended their participation in the study before the maximum four exposures to training and testing and their results are excluded from further analysis. In addition, eight participants failed to meet test criterion. For these participants, six (P2, P3, P6, P18, P23 and P24) achieved 100% accuracy on the baseline pairs, while accuracy for the two remaining participants (P4 and P6), ranged between 85% and 95%. Thus, irrespective of whether participants passed or failed, high accuracy was observed on the baseline pairs.

Table 2.2.
Individual data for participants that passed and failed Experiment 1.

Participant	Transitive Inference Training (trials to criterion)	Testing (% correct) Baseline	Endpoint	Non-endpoint
Pass(*n*=13)				
1	120	100	98	100
8	100	70	93	47
9	660	75	54	47
	20	95	71	48
	40	95	61	50
	20	90	79	53
10	120	90	64	53
	80	100	96	51
11	200	95	96	49
13	660	100	100	45
14	400	100	75	42
	20	100	57	45
	20	95	79	48
	20	100	86	52
15	260	85	86	57
17	120	75	96	53
19	160	95	57	47
	20	100	46	53
	20	100	71	61
	20	100	100	71
20	140	90	82	65
28	220	80	79	69
	40	100	100	83
29	380	75	79	75
	20	90	79	50
Fails(*n*=8)				
2*	220	95	54	33
	20	100	54	17
	20	100	64	17
	40	100	64	25
3*	160	100	36	33
	20	100	46	33
	20	100	71	33
	20	100	46	33
4*	120	85	64	33
	20	95	50	33
	20	100	75	33
	40	85	61	33
6*	180	90	54	33
	60	100	50	33
	40	90	39	33
	80	100	43	33
18*	100	100	57	33
	20	100	57	33

Table 2.2 *(cont.d)*	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
18*	20	100	43	33
	40	100	32	33
23*	200	60	64	33
	40	100	71	50
	20	90	54	33
	40	100	61	33
24*	260	95	75	25
	20	100	71	0
	20	100	86	17
	20	100	71	42
26*	60	100	61	33
	20	95	61	33
	20	100	57	33
	20	95	57	33

Note. Data is shown for the number of trials required to achieve transitive inference training criterion for participants that passed and failed the experimental task. The mean percent correct on the baseline, endpoint and non-endpoint test pairs, is also displayed for each participant during each exposure to testing. * refers to participants who failed to meet criterion at testing.

Endpoint Pairs. A graphical representation of the mean percent correct on the endpoint pairs for participants that passed and failed the experimental task can be seen in Figure 2.4. This graph shows that, accuracy on the endpoint pairs AF and BF was near perfect. Accuracy on the pairs AE, CF and DF was lower, but still above chance levels. However, accuracy on the AC and AD endpoint pairs was only just above chance levels. Average accuracy on the endpoint pairs ranged between 55% and 98%.

A summary of response accuracy for participants across each exposure to testing can be seen in Table 2.2. From this table, it can be seen that, for the thirteen participants that passed, three (P13, P19 and P28) achieved 100% accuracy on the endpoint pairs. For the other ten participants that passed (P1, P8, P9, P10, P11, P14, P15, P17, P20 and P29), accuracy on the endpoint pairs ranged between 79% and 98%. For the eight participants that failed to meet test criterion (P2, P3, P4, P6, P18, P23, P24 and P26), accuracy on the endpoint pairs ranged between 32% and 71%. Of these participants, three (P3, P6 and P18) performed below chance levels on the endpoint pairs. Thus, accuracy for participants that failed to meet test criterion was somewhat lower on the endpoint pairs, in comparison to those that passed.

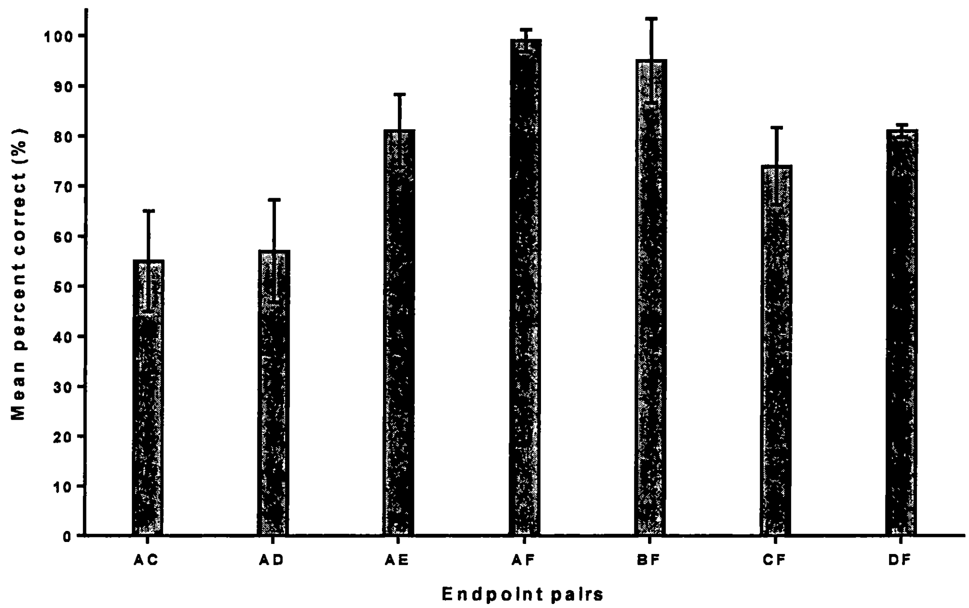


Figure 2.4. Experiment 1. Mean percent correct to the endpoint pairs for participants that passed and failed the experimental task. Error bars indicate standard errors.

Non-endpoint Pairs. A graphical representation of the mean percent correct on the non-endpoint pairs for participants that passed and failed the experimental task can be seen in Figure 2.5. This graph shows that, for participants that passed and failed, accuracy on the CE non-endpoint pair was high. In contrast, accuracy on the BD and BE pairs was only just above chance levels. Average accuracy on the non-endpoint pairs ranged between 57% and 87%.

A summary of response accuracy for participants across each exposure to testing can be seen in Table 2.2. From this table, it can be seen that, for the thirteen participants that passed, only one participant (P1) achieved 100% accuracy on the non-endpoint pairs. For the remaining twelve participants that passed (P8, P9, P10, P11, P13, P14, P15, P17, P19, P20, P28 and P29), accuracy on the non-endpoint pairs ranged between 45% and 83%. Indeed, a large number of participants that passed the experimental task displayed response accuracy only slightly above (P9, P10, P14, P15, P17 and P29), or below (P8, P11 and P13) chance levels on the non-endpoint pairs. For the eight participants that failed to meet test criterion (P2, P3, P4, P6, P18, P23, P24 and P26), accuracy on the non-endpoint pairs ranged from between 25% and 42%. Thus, all participants that failed to pass the experiment displayed response accuracy on the non-endpoint pairs that was below chance levels. In addition, accuracy on the non-endpoint pairs was, for a number of participants that passed, below, or just above chance levels.

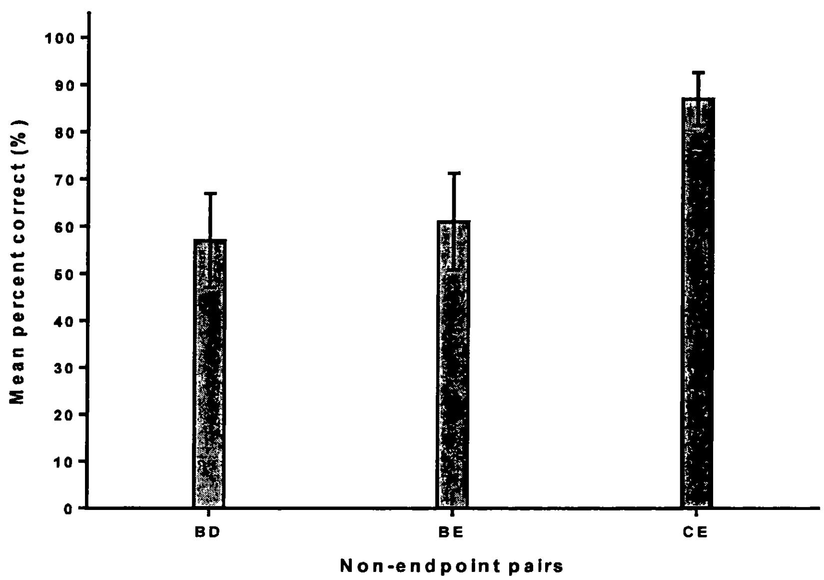


Figure 2.5. Experiment 1. Mean percent correct to the non-endpoint pairs for participants that passed and failed the experimental task. Error bars indicate standard errors.

Statistical Analyses

A McNemar test revealed that accuracy was significantly higher on the baseline pairs over the endpoint (p = .039) and non-endpoint pairs (p = .001). In addition, accuracy was significantly higher on the endpoint pairs over the non-endpoint pairs (p = .001).

In summary, only thirteen out of the twenty-nine participants that started Experiment 1 met criterion at testing. However, for these participants, high accuracy was observed on the baseline and endpoint pairs, but was somewhat lower on the non-endpoint pairs (see Table 2.2). For participants that failed to meet criterion at testing (see Table 2.2), high accuracy was maintained on the baseline pairs, but was somewhat lower on the endpoint pairs, and lower again on the non-endpoint pairs.

Experiment 1 had two primary aims, which were to incorporate pre-determined test mastery criterion into testing phases, and also, to explore the potential utility of repeated exposure to training and test phases in facilitating the emergence of TI. In Experiment 1, a minimum mean accuracy score of 80% (i.e., 48 correct out of 60) was in place during the test phase, which if not met, resulted in re-exposure to training and testing, for a maximum of three further times. In contrast, in a number of other TI studies, no test mastery criterion is employed, and participants are exposed to one test block in which the endpoint and non-endpoint pairs are presented alongside the baseline pairs (e.g., Greene et al., 2001; Lazareva & Wasserman, 2010; Libben & Titone, 2008; Moses et al., 2006, 2008). One potential problem with this method of testing is that if accurate responding to novel inferential test trials does not emerge immediately, then participants are not exposed to any form of intervention to remediate these weaknesses. However, in the current study, if the predicted transitive performances did not emerge immediately, participants were exposed to additional training and testing, in an attempt to facilitate the emergence of this behaviour. Findings revealed that for a number of participants (e.g., P2, P4, P14, P19 and P28), re-exposure to training and testing had the desired facilitative effects. That is, for these participants, additional exposure to training and test phases, allowed them to meet criterion at testing. Furthermore, for some of these participants (P19 and P28), performance accuracy on many of the endpoint and non-endpoint pairs, improved from their first to last exposure to testing (see Table 2.2). In addition, for P2, P4 and P14, performance accuracy improved on the endpoint, but not the non-endpoint pairs. However, it must be noted that, for participants that failed, repeated exposure to training and test phases, did not allow these participants to meet criterion at testing.

Furthermore, and as mentioned, findings from Experiment 1 revealed that for a number of participants that passed and failed, accuracy was low on the endpoint and non-endpoint pairs. More specifically, for a number of participants that passed (P10, P11, P13, P14 and P15), accuracy on the baseline and endpoint pairs was high but accuracy on the non-endpoint pairs was only at, or below chance levels. In addition, for participants that failed, high accuracy was maintained on the baseline pairs, but was at, or below chance levels on the endpoint and non-endpoint pairs. For example, for five of the participants that failed to meet test criterion (P2, P4, P23, P24 and P26), accuracy was above chance on the endpoint pairs, but below chance on the non-endpoint pairs. For the other three participants that failed (P3, P6 and P18), high accuracy was maintained on the baseline pairs, but was below chance on the endpoint and non-endpoint pairs. More specifically, for a number of participants that failed (P3, P4, P6, P18, P23 and P26), accuracy on the non-endpoint pairs was only 33%. Such findings would indicate that participants were responding correctly to only one of the non-endpoint pairs. That is, when participants were presented with, for example, the non-endpoint pairs BD, BE, and CE, participants may have selected the correct discriminative stimulus to only one of these pairs. Indeed, results demonstrated that high accuracy was observed on the CE non-endpoint pair, but was somewhat lower on the BD and BE pairs. Thus, participants may have incorrectly selected D in the BD pair, and E in the BE pair. In turn, such findings appear to suggest that for a number of participants, the incorrect discriminative stimulus (S-) exerted control over responding.

Experiment 1 also compared performance accuracy on the endpoint and non-endpoint test pairs. Findings from Experiment 1 revealed that accuracy was significantly higher on the endpoint test pairs in comparison to the non-endpoint test pairs. In turn, such findings correspond to proposals by associative-learning, and reinforcement-based accounts of TI, that more accurate responding is often observed on test pairs containing end terms (e.g., AC, AD, BF and CF) in comparison to those that do not contain end terms (e.g., BD, BE and CE; Bryant & Trabasso, 1971; Frank et al., 2003; van Elzakker et al., 2003; Wynne et al., 1992). Indeed, in Experiment 1, all the endpoint pairs contain an end term (A and F), while none of the non-endpoint pairs did (e.g., BD and BE). Thus, findings from Experiment 1 would appear to suggest that participants employed lower-level associative learning principles to solve the TI task. However, and as mentioned in previous sections (Chapter 1), considerable debate exists as to whether humans employ higher-order strategies, such as awareness, or lower-level associative-learning strategies to solve the TI task. Experiment 1 did not examine the role of awareness on inferential performances, and thus, it is questionable as to whether such differences in performances on the endpoint and non-endpoint test pairs would be observed, if participants were aware of the underlying stimulus hierarchy. Indeed, previous findings suggest that awareness of the stimulus hierarchy is beneficial for successful inferential performances (e.g., Moses et al., 2006, 2008), and thus, Experiment 2 sought to explore this.

Experiment 2

Experiment 2 sought to examine the role of awareness on the emergence of TI, and is partly based on studies by Frank et al. (2005) and Greene et al. (2001). For a number of participants in Experiment 1, accuracy was below chance levels on the endpoint and non-endpoint pairs, and it was questioned as to whether awareness of the underlying stimulus hierarchy may lead to more accurate inferential test performances. However, the method in which awareness is currently assessed in TI differs between studies. For instance, Frank et al. (2005) exposed participants to testing with three inferential pairs (BD, BE and CE) along with the endpoint pair AF, following training on five adjacent stimulus pairs (AB, BC, CD, DE and EF). In order to examine the role of awareness, Frank et al. (2005) classified participants as Aware on the basis of responses to post-experimental questionnaires. Thus, participants were considered to be Aware if they could determine the underlying stimulus hierarchy, whereas participants who failed to do so were classified as Unaware. Findings demonstrated that participants that were classified as Unaware of the stimulus hierarchy reliably chose B over E in the BE pair, but performance on the BD and CE pairs was at chance levels (for a theoretical account of superior BE test performances for Unaware participants, see Frank et al., 2003; van Elzakker et al., 2003; Wynne, 1995, 1998). In contrast, Aware participants displayed near-perfect accuracy on all test pairs, and thus, Frank et al. (2005) proposed that participants that were unaware of the hierarchy were capable of displaying inferential responding in the absence of awareness. Similarly, Moses et al. (2006, 2008) examined the role of awareness in a similar way to Frank et al. (2005). However, Moses et al. (2006, 2008) found that performance accuracy on three inferential test pairs (BD, CE and

BE) was significantly higher for participants that were Aware of the stimulus hierarchy in comparison to those that were Unaware. In addition, participants that were Unaware of the hierarchy performed at chance levels on the non-endpoint pairs as well as some premise and endpoint pairs (Experiment 1). Moreover, post-experimental measures of awareness were correlated with accurate performances on the inferential pairs for the Aware, but not, Unaware group.

In contrast, Greene et al. (2001) examined transitive responding for an Informed and Uninformed group. In contrast to the Frank et al. (2005) study, one group of participants were given additional instructions at the start of the experiment that the stimuli could be arranged in a hierarchy (Informed), while a second group were not given this information (Uninformed; see also Lazareva & Wasserman, 2010). Findings demonstrated that awareness of the hierarchy led to faster acquisition in the ability to make inferential judgements and that participants in the Informed group displayed higher levels of accuracy than those in the Uninformed group on the BD pair. However, Uninformed participants also performed significantly above chance on the BD test pair, and successful inferential performance was not correlated with a post-experimental measure of awareness. Furthermore, Greene et al. found that when performance on BD was near-perfect, task awareness was not high. This in turn led Greene et al. (2001) to propose that although explicit awareness of the stimulus hierarchy is sufficient for TI, it may not be necessary.

Experiment 2 therefore sought to further explore the role of awareness on TI, and similar to Greene et al. (2001), one group of participants received explicit instructions at the start of the experiment that the stimuli could be arranged in a hierarchy (Informed), while a second group did not (Uninformed). In addition, in order to address a limitation in the Greene et al. (2001) study, and on the basis of proposals by Lazareva and Wasserman (2010), Experiment 2 sought to further determine the specific role of pre-experimental instructions on inferential performances by examining responses to post-experimental questionnaires for the Informed and Uninformed groups. For example, Lazareva and Wasserman (2010) reported that providing participants with additional instructions at the start of the experiment does not guarantee awareness at the end of the study. In turn, such findings question the specific role of awareness on the emergence of TI, and thus, in Experiment 2, participants in the Informed and Uninformed groups were required to complete post-experimental awareness questionnaires. In addition, similar to Experiment 1, mastery criterion was employed during test phases in Experiment 2, and participants were exposed to additional training and test phases, if they initially failed to meet test criterion.

Method

Participants

Forty-three students, 20 male and 23 female, ranging in age from 18 to 33 years (M_{age} = 20.51, *SD* = 2.59) were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the study, and were randomly assigned to the Informed or Uninformed groups at the start of the experiment.

Procedure

The procedure for Experiment 2 was identical to Experiment 1, with the exception that, prior to training, half of the participants were randomly assigned to an Informed group, in which they were told that there was an underlying hierarchy between the stimuli. The second group, the Uninformed group, did not receive this additional information. The instructions presented to the Uninformed group were the same as those presented in Experiment 1, while the instructions for the Informed group, were as follows:

> You will be presented with two images, centred on the right- and left-hand side of the computer screen. Your task is to learn to select the correct image. There is an underlying hierarchy among the images. Your task is to learn this hierarchy. To select the image on the left, press the marked key on the left of the keyboard. To select the image on the right, press the marked key on the right of the keyboard. Sometimes the computer will give you feedback, and at other times it will not. It is possible to get all of the tasks without feedback correct, by paying close attention to the tasks

with feedback. The computer will tell you when this phase of the experiment is finished.

Thus, the only difference between the instructions presented to the Informed and Uninformed groups were that participants in the Informed group were instructed that "There is an underlying hierarchy among the images. Your task is to learn this hierarchy". In addition, mastery criterion for *Phase 1: Transitive Inference Training* differed from Experiment 1. Participants in Experiment 2 were now required to achieve a minimum mean accuracy of 90% (i.e., 18 out of 20 correct responses) on the baseline pairs during transitive inference training. If participants failed to do so, training blocks were repeated until participants achieved this criterion. Mastery criterion for *Phase 2: Testing* remained the same.

Results and Discussion

Of the forty-three participants that started Experiment 2, four participants from the Uninformed group terminated their participation in the experiment before the maximum four exposures to testing, and their data is therefore excluded from further analyses. For the remaining thirty-nine participants, four from the Informed group and seven from the Uninformed group failed to meet criterion following four exposures to testing. On the other hand, a total of twenty-eight participants (Informed: 16; Uninformed: 12) passed the experimental task and required between 1 and 4 exposures to testing to do so (Informed: M = 1.63; SD = .96; Uninformed: M = 1.92; SD = .99). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Note. Due to experimenter error, it was not possible to examine the post-experimental questionnaires for either the Informed or Uninformed group of participants.

Training trials to criterion

In order to meet criterion during transitive inference training, participants in the Informed group required between 20 and 680 trials (M = 146.25, SD = 181.80), while those in the Uninformed group required between 20 and 620 trials (M =

116.67 *SD* = 169.03). An independent t-test revealed there were no significant differences between the Informed and Uninformed group on the mean number of training trials required to reach criterion ($t(26) = .44$, $p = .66$; see Tables 2.3 (Informed) and 2.4 (Uninformed), for a summary of participants' performance accuracy on each exposure to training and testing).

Accuracy: Testing

Baseline pairs. A graphical representation of the mean percent correct on the baseline pairs for participants in the Informed and Uninformed groups that passed and failed the experimental task can be seen in Figure 2.6. This graph shows that high accuracy was maintained on the baseline pairs, and was comparable between the groups. Average accuracy for the Informed group on the baseline pairs ranged between 88% and 95%, while accuracy ranged between 88% and 96% for the Uninformed group.

A summary of response accuracy for participants in the Informed group, across each exposure to testing can be seen in Table 2.3. From this table, it can be seen that, for the sixteen participants in the Informed group that passed, eight (P3, P4, P6, P7, P12, P13, P14 and P18) achieved 100% accuracy on the baseline pairs. For the remaining eight participants in the Informed group that passed (P2, P5, P9, P10, P11, P15, P19 and P20), accuracy on the baseline pairs ranged between 60% and 95% correct. A total of four participants in the Informed group failed to meet test criterion (P1, P8, P16 and P17), but displayed accuracy ranging between 80% and 95% on the baseline pairs.

Table 2.3.

Individual data for participants in the Informed group that passed and failed Experiment 2.

	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
Pass (n=16)				
2	180	95	100	100
3	60	100	100	100
4	100	100	100	100
5	260	85	96	50
6	680	100	100	100
7	100	85	57	42
	40	100	93	75
9	460	80	89	58
10	160	70	68	17
	40	80	75	42
	60	95	100	100
11	80	60	96	92
12	100	100	100	100
13	140	95	43	75
	60	95	68	33
	20	100	96	67
14	140	65	79	75
	40	100	100	100
15	140	90	79	100
18	140	70	93	58
	20	100	100	92
19	120	95	71	58
	20	95	57	50
	20	100	75	25
	20	90	96	100
20	80	95	96	100
Fails(n=4)				
1*	140	80	64	17
	40	90	39	25
	20	100	43	33
	20	90	64	42
8*	80	75	43	33
	20	90	54	33
	20	100	57	33
	20	95	57	33
16*	280	90	68	50
	20	65	57	33
	20	60	50	42
	40	85	54	33

Table 2.3 (*cont.d*) Participant	Transitive Inference Training (trials to criterion)	Testing (% correct) Baseline	Endpoint	Non-endpoint
17*	360	70	54	100
	60	75	61	100
	40	90	68	75
	40	80	57	67

Note. Data is shown for the number of trials required to achieve transitive inference training criterion for participants that passed and failed the experimental task. The mean percent correct on the baseline, endpoint and non-endpoint test pairs, is also displayed for each participant during each exposure to testing. * refers to participants who failed to meet criterion at testing.

A summary of response accuracy for participants in the Uninformed group, across each exposure to testing, can be seen in Table 2.4. From this table, it can be seen that, for the twelve participants in the Uninformed group that passed, five (P7, P9, P12, P16 and P19) achieved 100% accuracy on the baseline pairs. For the remaining seven participants in the Uninformed group that passed (P1, P4, P5, P8, P13, P15 and P17), accuracy on the baseline pairs ranged between 70% and 90% correct. A total of seven participants in the Uninformed group failed to meet test criterion. In addition, one participant in the Uninformed group (P18) failed to complete the experimental task, and his/her results are excluded from this analysis. With respect to the participants in the Uninformed group that failed to meet test criterion (P2, P3, P6, P10, P11, P14 and P20), accuracy on the baseline pairs ranged between 50% and 100% correct, with only one of these participants (P20) performing at chance levels on the baseline pairs. In summary, high accuracy was maintained for both groups on the baseline pairs, and accuracy was comparable between the groups.

Table 2.4.
Individual data for participants in the Uninformed group that passed and failed Experiment 2.

	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
Pass(*n*=12)				
1	100	75	100	92
4	620	70	82	92
5	80	80	89	100
7	800	65	75	67
	20	100	86	92
8	160	85	89	83
9	640	90	54	67
	40	100	71	83
12	160	85	71	25
	20	100	75	25
	20	100	61	33
	20	100	100	25
13	200	90	96	100
15	860	85	75	75
	20	95	50	50
	20	90	100	100
16	40	80	86	42
	20	90	57	25
	20	100	100	100
17	260	80	89	42
19	100	45	61	58
	80	100	100	100
Fails (*n*=8)				
2*	200	95	57	33
	20	100	57	33
	20	100	71	33
	20	95	50	33
3*	100	80	75	42
	20	95	61	33
	40	95	57	33
	20	100	54	33
6*	120	95	57	33
	20	100	57	33
	20	100	57	33
	20	95	57	33
10*	80	85	64	58
	40	90	71	42
	20	75	75	33
	40	85	71	42
11*	160	80	68	58
	20	75	46	42
	20	95	68	33
	20	95	79	42

Table 2.4 *(cont.d)*	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
14*	380	50	36	58
	60	95	39	33
	20	100	39	33
	20	100	43	33
18**	220	75	75	33
	40	90	69	58
	80	90	64	42
20*	180	95	64	42
	20	100	46	17
	40	100	46	42
	20	50	57	0

Note. Data is shown for the number of trials required to achieve transitive inference training criterion for participants that passed and failed the experimental task. The mean percent correct on the baseline, endpoint and non-endpoint test pairs, is also displayed for each participant during each exposure to testing. * refers to participants who failed to meet criterion at testing. ** refers to participants that terminated their participation in the experiment.

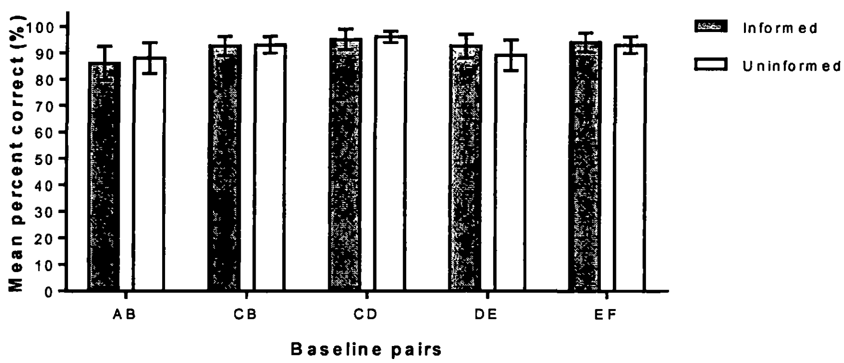


Figure 2.6. Experiment 2. Mean percent correct to the baseline pairs for participants in the Informed and Uninformed groups that passed and failed the experiment. Error bars indicate standard errors.

Endpoint pairs. A graphical representation of the mean percent correct on the endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experimental task can be seen in Figure 2.7. This graph shows that, for the most part, high accuracy was observed for both groups on the endpoint pairs. However, accuracy on the endpoint pairs, AC, AD and BF were considerably higher for participants in the Informed group relative to the Uninformed group. Average accuracy for the Informed group on the endpoint pairs ranged between 79% and 96%, while accuracy ranged between 60% and 96% for the Uninformed group.

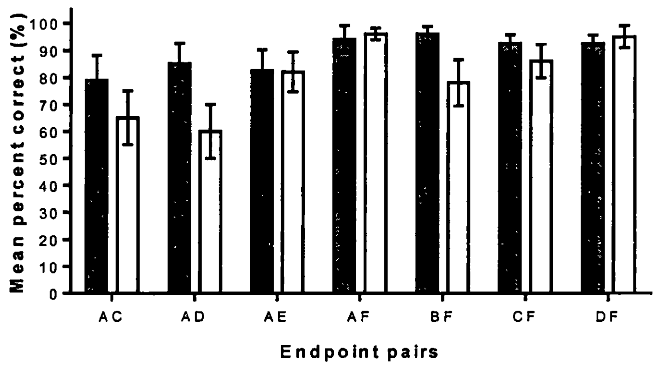


Figure 2.7. Experiment 2. Mean percent correct to the endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experiment. Error bars indicate standard errors.

A summary of response accuracy for participants in the Informed group, across each exposure to testing can be seen in Table 2.3. From this table, it can be seen that, for the sixteen participants in the Informed group that passed, eight (P2, P3, P4, P6, P10, P12, P14 and P18) achieved 100% accuracy on the endpoint pairs. For the remaining eight participants in the Informed group that passed (P5, P7, P9, P11, P13, P15, P19 and P20), accuracy on the endpoint pairs ranged between 79% and 96% correct. For the four participants in the Informed group that failed to meet test criterion, accuracy on the endpoint pairs ranged between 54% and 64%. (P1, P8, P16 and P17).

A summary of response accuracy for participants in the Uninformed group, across each exposure to testing can be seen in Table 2.4. From this table, it can be seen that, for the twelve participants in the Uninformed group that passed, five (P2, P12, P15, P16 and P19) achieved 100% accuracy on the endpoint pairs. For the remaining seven participants in the Uninformed group that passed (P4, P5, P7, P8, P9, P13 and P17), accuracy on the endpoint pairs ranged between 82% and 96% correct. For the seven participants in the Uninformed group that failed to meet test criterion (P2, P3, P6, P10, P11, P14 and P20), accuracy on the endpoint pairs ranged between 43% and 79%. However, only two of these participants (P2 and P12) performed at, or below chance levels on the endpoint pairs. In summary, for the most part, accuracy on the endpoint pairs was higher for participants in the Informed group, relative to the Uninformed group.

Non-endpoint pairs. A graphical representation of the mean percent correct on the non-endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experimental task can be seen in Figure 2.8. This graph shows that, for the most part, accuracy on the non-endpoint pairs was higher for participants in the Informed group in comparison to those in the Uninformed group. More specifically, accuracy for the Uninformed group on the BD pair was 54%, while accuracy for the Informed group on this same test pair was 81%. Similarly, accuracy for the Uninformed group on the BE pair was 63%, while accuracy for the Informed group on this same test pair was 76%. In contrast, accuracy on the non-endpoint pair, CE, was high for both groups (Informed: 91%; Uninformed: 86%). In summary, average accuracy for the Informed group on the non-endpoint pairs ranged between 76% and 91%, while accuracy ranged between 54% and 86% for the Uninformed group.

A summary of response accuracy for participants in the Informed group, across each exposure to testing, can be seen in Table 2.3. From this table, it can be seen that, for the sixteen participants in the Informed group that passed, ten (P2, P3, P4, P6, P10, P12, P14, P15, P19 and P20) achieved 100% accuracy on the non-endpoint pairs. For the remaining six participants in the Informed group that passed (P5, P7, P9, P11, P13 and P18), accuracy on the non-endpoint pairs ranged between 50% and 75% correct, with only one participant (P5) performing at chance levels. For the four participants in the Informed group that failed to meet test criterion (P1, P8, P16 and P17), accuracy on the non-endpoint pairs ranged between 33% and 67%.

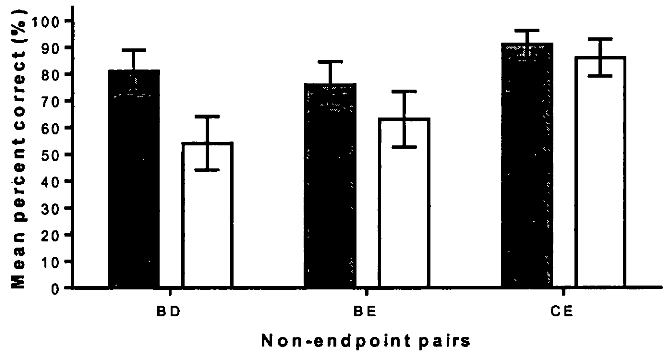


Figure 2.8. Experiment 2. Mean percent correct to the non-endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experiment. Error bars indicate standard errors.

A summary of response accuracy for participants in the Uninformed group, across each exposure to testing, can be seen in Table 2.4. From this table, it can be seen that, of the twelve participants in the Uninformed group that passed, five achieved 100% accuracy on the non-endpoint pairs (P5, P13, P15, P16 and P19). For the remaining seven participants in the Uninformed group that passed (P1, P4, P7, P8, P9, P13 and P17), accuracy on the non-endpoint pairs ranged between 25% and 92% correct. It must be noted that, two of these participants (P12 and P17) performed below chance on the non-endpoint pairs. For the seven participants in the Uninformed group that failed to meet test criterion (P2, P3, P6, P10, P11, P14 and P20), accuracy on the non-endpoint pairs ranged between 0% and 42% correct. In summary, participants in the Informed group outperformed those in the Uninformed group on the non-endpoint pairs BD and BE, but performances were comparable between the groups on the non-endpoint pair, CE.

Statistical Analyses

A McNemar test revealed no significant differences between accuracy scores for participants in the Informed group on the baseline and endpoint pairs (p = .219). Similarly, no significant differences were observed between accuracy scores for the Informed group on the endpoint and non-endpoint pairs (p = .375). However, there was a significant difference in accuracy scores between the baseline and non-endpoint pairs for participants in the Informed group (p = .039). Thus, participants in the Informed group displayed significantly higher levels of accuracy on the baseline pairs over the non-endpoint pairs.

A McNemar test revealed no significant differences in accuracy scores for participants in the Uninformed group between the baseline and endpoint (p = .180), non-endpoint pairs (p = .109), and the endpoint and non-endpoint pairs (p = 1.000).

In summary, sixteen participants from the Informed group, and twelve from the Uninformed group, met criterion at testing. For both groups, high accuracy was maintained on the baseline pairs, with little differences observed between the groups. However, differences between the Informed and Uninformed groups became apparent when accuracy on the endpoint and non-endpoint pairs was examined. For instance, high accuracy was observed for the Informed group on the endpoint (79%-96%) and non-endpoint pairs (76%-91%), but was considerably lower for the Uninformed group on these same test pairs (Endpoint: 60%-96%; Non-endpoint: 54%-86%). Participants in the Uninformed group only performed just above chance levels on the BD non-endpoint pair (54%). However, slightly higher accuracy was observed for the Uninformed group on the BE non-endpoint pair (63%). In contrast, accuracy on the non-endpoint pair CE was comparable between the groups (Informed: 91%; Uninformed: 86%).

Findings from Experiment 2 would appear to suggest that awareness of the stimulus hierarchy has a facilitative effect on inferential performances at test. These findings are in contrast to Greene et al.'s (2001), who found that although accuracy was higher for participants in the Informed group on the BD inferential pair, participants in the Uninformed group performed significantly above chance on the BD pair. However, an important difference between the current study and Greene et al.'s was that participants in the current study were exposed to training and testing with a 6-term series, whereas participants in the Greene et al. (2001) study were exposed to training and testing with a 5-term series. Thus, it is questionable as to whether participants in the Uninformed group in the Greene et al. (2001) study would have performed above chance levels on the additional non-endpoint pairs, BE and CE, presented in the current study. Furthermore, and in contrast to the current study, in the Greene et al. (2001) study, participants were exposed to BD probe trials at the end of each training block (Experiment 2B), and accuracy on these probe trials was assessed. Thus, in the Greene et al. (2001) study, if participants made a minimum of seven out of eight correct responses on the BD probe trials at the end of a given training block, the training part of the experiment ended, and participants were required to complete a post-experimental awareness questionnaire. Greene et al. (2001) included these probe trials in order to determine whether awareness precedes performance. Thus, presenting BD probe trials at the end of training blocks may have resulted in more accurate transitive performances for the Uninformed group in the Greene et al. (2001) study.

An interesting finding from Experiment 2 was that, accuracy for participants in the Informed and Uninformed groups that failed to meet criterion at testing (Informed: P1, P8, P16 and P17; Uninformed: P2, P3, P6, P10, P11, P14, P16 and P20), was high on the baseline pairs, but was somewhat lower on the endpoint pairs, and lower again on the non-endpoint pairs (Informed: Endpoint pairs: 54%-64%; Non-endpoint pairs: 33%-62%; Uninformed: Endpoint pairs: 50%-79%; Non-endpoint pairs: 0%-42%). More specifically, and similar to findings from Experiment 1, accuracy for a number of participants was only 33% on the non-endpoint pairs (Informed: P8 and P16; Uninformed: P2, P3, P6 and P14). Thus, similar to proposals in the discussion of Experiment 1, such findings suggest that participants may have responded correctly to only one of the non-endpoint pairs. Indeed, high accuracy was only observed on the non-endpoint pair CE, irrespective of whether participants passed or failed, and in turn, suggests, that participants may have encountered difficulties on the non-endpoint pairs, BD and BE. Therefore, in order to determine whether more accurate responding on these two non-endpoint pairs may be achieved, and similar to Greene et al. (2001), Experiment 3 sought to examine whether presenting BD probe trials throughout training blocks would have a facilitative effect on inferential responding.

In addition, due to experimenter error with the post-experiment questionnaires in Experiment 2, Experiment 3 sought to determine the influence of pre-experimental instructions on post-experimental measures of awareness, by examining post-experimental levels of awareness for the Informed and Uninformed groups.

Experiment 3

As mentioned, Experiment 3 sought to extend the findings of Experiment 2. Performances on novel inferential test trials were compared between an Informed and Uninformed group of participants. In addition, both groups were required to complete a post-experimental questionnaire to assess, whether awareness of the stimulus hierarchy was correlated with test performances for both groups. Furthermore, Experiment 3 sought to determine whether presenting probe trials for the non-endpoint pair, BD, throughout training blocks, facilitates inferential responding at test. Although most studies have presented non-adjacent stimulus pairs during the test phase only, Greene et al. (2001) suggested that presenting the critical BD probe trials at the end of training blocks, might more readily identify the time course of the emergence of awareness and successful inferential performance. In Experiment 3 of the current study, four unreinforced (i.e., no feedback was provided on these trials) probe trials were included throughout all training blocks, and accuracy on the BD probe trial was not assessed. Thus, the current study was concerned with whether probing for BD performance throughout training would facilitate accurate responding on other inferential pairs at test (e.g., BE and CE).

Method

Participants

Forty students, eight male and thirty-two female, ranging in age from 18 to 49 years (M_{age} = 22.28, SD = 6.42) were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit, or paid £6, on completion of the study, and were randomly assigned to the Informed or Uninformed group at the start of the experiment.

Procedure

The procedure for Experiment 3 was identical to Experiment 2, with the exception of transitive inference training, in that participants were now presented with probe trials for the non-endpoint test pair, BD.

Transitive Inference Training. Similar to Experiments 1 and 2, participants were exposed to training with five adjacent stimulus pairs (A+B-, B+C-, C+D-, D+E-and E+F-). However, in addition, participants were presented with four probe trials, in which the BD test pair was presented in the absence of reinforcement. Thus, participants were exposed to a total of twenty training trials and four BD probe trials during training, and were required to achieve a minimum mean of 90% (i.e., 18 out of 20 correct responses) on the five baseline pairs (AB, BC, CD, DE and EF). Accuracy on the BD probe pair did not affect mastery criterion during transitive inference training, and no other inferential (non-endpoint) probe trials were presented.

Testing. This test phase was identical to Experiments 1 and 2.

Post-experimental Awareness Questionnaire. Upon completion of the experiment, participants were provided with a post-experimental questionnaire to assess their awareness of the TI task and the test trials that did not contain endpoints (i.e., CE, BD and BE). In addition, the questionnaire sought to determine what strategies, if any, participants used to respond to novel test pairs (see Appendix 4). The awareness rating scale for each question ranged from 0 to 2, with a score of 0 corresponding to no awareness of the 6-term linear series, 1 corresponding to some evidence of awareness, and 2 corresponding to definite indications of awareness.

Results and Discussion

Of the forty participants that started Experiment 3, four from the Uninformed group terminated their participation in the experiment before the maximum four exposures to testing, and their data is therefore excluded from further analyses. For the remaining thirty-six participants, four from the Informed group and three from the Uninformed group failed to meet criterion at testing. On the other hand, a total of twenty-nine participants (Informed: 16; Uninformed: 13) passed the experimental task and required between 1 and 3 exposures to testing to do so (Informed: $M = 1.38$; $SD = .50$; Uninformed: $M = 2.09$; $SD = .76$). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Training trials to criterion

In order to meet criterion during the transitive inference training, participants in the Informed group required between 24 and 360 trials ($M = 199.50$, $SD =$

114.17), while those in the Uninformed group required between 24 and 432 trials (*M* = 276.92, *SD* = 143.87). An independent t-test revealed there were no significant differences between the Informed and Uninformed groups on the mean number of training trials required to reach criterion ($t(27) = -1.62$, $p = .12$; see Tables 2.5 (Informed) and 2.6 (Uninformed) for a summary of participants' performance accuracy on each exposure to training and testing).

Accuracy: Testing

Baseline pairs. A graphical representation of the mean percent correct on the baseline pairs for participants in the Informed and Uninformed groups that passed and failed the experimental task can be seen in Figure 2.9. This graph shows that high accuracy was maintained on the baseline pairs, and was comparable between the Informed and Uninformed groups. Average accuracy on the baseline pairs ranged between 86% and 99% for participants in the Informed group, and between 89% and 100% for participants in the Uninformed group.

A summary of response accuracy for participants in the Informed group across each exposure to testing can be seen in Table 2.5. From this table, it can be seen that, for the sixteen participants in the Informed group that passed, six (P1, P3, P7, P8, P13 and P17) achieved 100% accuracy on the baseline pairs. For the remaining ten participants in the Informed group that passed (P2, P4, P5, P6, P9, P11, P12, P14, P16 and P20), accuracy on the baseline pairs ranged between 75% and 95%. A total of four participants in the Informed group (P10, P15, P18 and P19) failed to meet criterion at testing. For these participants, accuracy on the baseline pairs ranged between 75% and 100% correct.

Table 2.5.
Individual data for participants in the Informed group that passed and failed Experiment 3.

	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
Pass(*n*=16)				
1	96	100	100	100
2	240	80	93	83
3	72	100	96	92
4	48	90	100	100
5	312	65	82	50
	144	75	93	67
6	96	70	68	0
	48	95	82	50
7	288	95	61	25
	24	100	100	100
8	192	100	89	58
9	240	75	79	83
	72	80	82	83
11	216	80	89	100
12	168	60	61	58
	24	95	96	100
13	360	100	100	100
14	192	90	75	100
16	168	80	96	100
17	96	100	61	42
	24	100	71	67
20	72	90	96	100
Fails (*n*=4)				
10*	144	90	64	42
	24	85	61	67
	24	95	57	67
	24	95	39	67
15*	96	95	50	33
	24	70	50	42
	24	100	46	33
	24	75	57	25
18*	240	100	61	33
	48	95	46	33
	24	100	61	42
	24	100	54	33
19*	192	85	18	0
	48	100	68	0
	24	100	82	0
	24	100	82	0

Note. Data is shown for the number of trials required to achieve transitive inference training criterion for participants that passed and failed the experimental task. The mean percent correct on the baseline, endpoint and non-endpoint test pairs, is also displayed for each participant during each exposure to testing. * refers to participants who failed to meet criterion at testing.

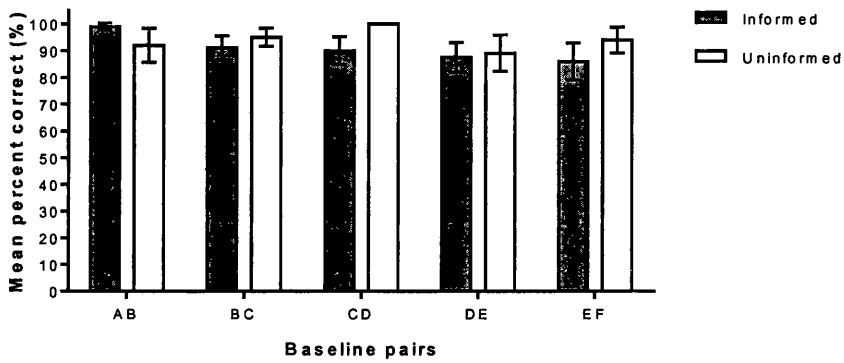


Figure 2.9. Experiment 3. Mean percent correct to the baseline pairs for participants in the Informed and Uninformed groups that passed and failed the experiment. Error bars indicate standard errors.

A summary of response accuracy for participants in the Uninformed group across each exposure to testing can be seen in Table 2.6. From this table, it can be seen that, for the thirteen participants in the Uninformed group that passed, nine (P1, P3, P8, P9, P11, P12, P13, P15 and P16) achieved 100% accuracy on the baseline pairs. For the remaining four participants in the Uninformed group that passed (P17, P18, P19 and P20), accuracy on the baseline pairs ranged between 75% and 90% correct. A total of three participants in the Uninformed group (P5, P6 and P10) failed to meet criterion at testing. In addition, a further four participants (P2, P4, P7 and P14) failed to complete the experimental task and their results will be excluded from further analysis. With respect to the participants in the Uninformed group that failed to meet test criterion, accuracy on the baseline pairs ranged between 90% and 100% correct. In summary, high accuracy was observed for the Informed and Uninformed groups on the baseline pairs, irrespective of whether participants passed or failed.

Table 2.6.
Individual data for participants in the Uninformed group that passed and failed Experiment 3.

	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
Pass(*n*=13)				
1	96	100	100	100
3	288	100	96	92
8	120	70	68	33
	48	100	100	100
9	168	100	43	33
	24	100	64	58
	24	100	100	58
11	336	90	79	33
	24	95	82	33
	48	100	100	100
12	192	95	71	33
	24	100	86	33
13	192	75	46	83
	24	100	86	100
15	264	75	39	33
	48	80	54	33
	24	100	86	100
16	240	75	71	42
	48	100	86	25
	24	100	100	100
17	240	65	64	42
	432	75	96	92
18	216	80	61	25
	72	90	75	75
19	192	80	86	92
20	144	72	79	33
	48	75	89	67
Fails (*n*=7)				
2**	192	100	39	75
	48	95	44	58
	96			
4**	720			
5*	72	100	61	0
	48	100	43	0
	48	90	39	25
	24	100	43	33
6*	264	80	61	33
	24	80	39	33
	24	85	54	33
	48	90	57	33

Table 2.6 (*cont.d*)	Transitive Inference Training	Testing (% correct)		
Participant	(trials to criterion)	Baseline	Endpoint	Non-endpoint
7**	640			
10*	120	85	61	25
	48	100	50	33
	24	95	46	33
	24	95	57	33
14**	942			

Note. Data is shown for the number of trials required to achieve transitive inference training criterion for participants that passed and failed the experimental task. The mean percent correct on the baseline, endpoint and non-endpoint test pairs, is also displayed for each participant during each exposure to testing. * refers to participants who failed to meet criterion at testing. ** refers to participants that terminated their participation in the experiment.

Endpoint pairs. A graphical representation of the mean percent correct on the endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experimental task can be seen in Figure 2.10. This graph shows that high accuracy was observed on the endpoint pairs for both the Informed and Uninformed groups. However, for the most part, slightly higher accuracy was observed for the Informed group on all endpoint pairs. In addition, higher accuracy was observed for both groups on the endpoint pairs AD, AE and AF, in comparison to the other endpoint pairs (AC, BF, CF and DF). Average accuracy on the endpoint pairs ranged between 75% and 97% for participants in the Informed group, and between 76% and 95% for participants in the Uninformed group.

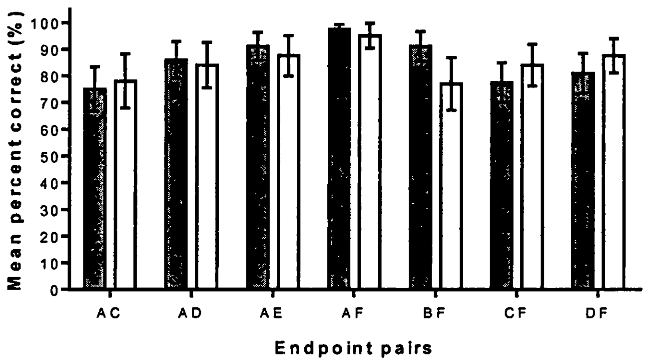


Figure 2.10. Experiment 3. Mean percent correct to the endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experiment. Error bars indicate standard errors.

A summary of response accuracy for participants in the Informed group across each exposure to testing can be seen in Table 2.5. From this table, it can be seen that, for the sixteen participants in the Informed group that passed, four (P1, P4, P7 and P13) achieved 100% accuracy on the endpoint pairs. For the remaining twelve participants in the Informed group that passed (P2, P3, P5, P6, P8, P9, P11, P12, P14, P16, P17 and P20), accuracy on the endpoint pairs ranged between 71% and 96%. For the four participants in the Informed group that failed to meet test criterion (P10, P15, P18 and P19), accuracy on the endpoint pairs ranged between 39% and 82%, with only one participant (P10) responding below chance levels.

A summary of response accuracy for participants in the Uninformed group across each exposure to testing can be seen in Table 2.6. From this table, it can be seen that, for the thirteen participants in the Uninformed group that passed, five (P1, P8, P9, P11 and P16) achieved 100% accuracy on the endpoint pairs. The remaining eight participants in the Uninformed group that passed (P3, P12, P13, P15, P17, P18, P19 and P20), made between 75% and 96% correct responses on the endpoint pairs. For the three participants in the Uninformed group that failed to meet test criterion (P5, P6 and P10), accuracy on the endpoint pairs ranged between 43% and 57% correct. In summary, high accuracy was observed on the endpoint pairs for both the Informed and Uninformed groups, and a slight performance advantage was noted for the Informed group on these test pairs.

Non-endpoint pairs. A graphical representation of the mean percent correct on the non-endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experimental task can be seen in Figure 2.11. This graph shows that accuracy on the non-endpoint pairs BD and BE was higher for participants in the Informed group in comparison to those in the Uninformed group. However, participants in the Uninformed group performed above chance levels on the non-endpoint pairs, BD (64%) and BE (71%). In addition, high accuracy was observed on the CE non-endpoint pair, and was slightly higher for the Uninformed group (Informed: 85%; Uninformed: 91%). Average accuracy on the non-endpoint pairs ranged between 72.5% and 85% for participants in the Informed group, and between 64% and 90% for participants in the Uninformed group.

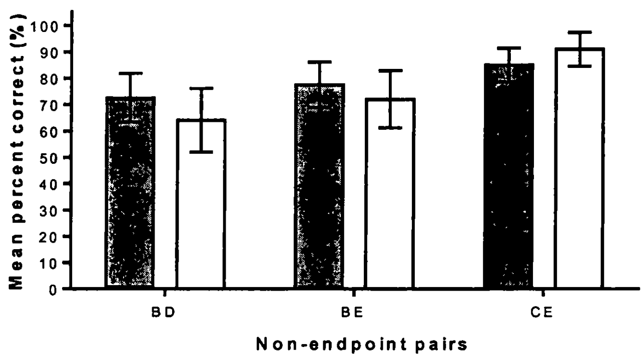


Figure 2.11. Experiment 3. Mean percent correct to the non-endpoint pairs for participants in the Informed and Uninformed groups that passed and failed the experiment. Error bars indicate standard errors.

A summary of response accuracy for participants in the Informed group across each exposure to testing can be seen in Table 2.5. From this table, it can be seen that, for the sixteen participants in the Informed group that passed, nine (P1, P4, P7, P11, P12, P13, P14, P16, and P20) achieved 100% accuracy on the non-endpoint pairs. For the remaining seven participants in the Informed group that passed (P2, P3, P5, P6, P8, P9 and P17), accuracy on the non-endpoint pairs ranged between 50% and 92% correct, with only one participant (P6) performing at chance levels. For participants in the Informed group that failed to meet test criterion (P10, P15, P18 and P19), accuracy on the non-endpoint pairs ranged between 0% and 67% correct.

Indeed, P10 was the only participant to perform above chance on the non-endpoint pairs.

A summary of response accuracy for participants in the Uninformed group across each exposure to testing can be seen in Table 2.6. From this table, it can be seen that, for the thirteen participants in the Uninformed group that passed the experiment, six (P1, P8, P11, P13, P15 and P16) achieved 100% accuracy on the non-endpoint pairs. For the remaining seven participants in the Uninformed group that passed (P3, P9, P12, P17, P18, P19 and P20), accuracy on the non-endpoint pairs ranged between 33% and 92% correct, with only one participant (P12) performing below chance levels. For participants in the Uninformed group that failed to meet test criterion (P5, P6 and P10), all three achieved accuracy of 33% on the non-endpoint pairs. In summary, higher accuracy was observed for participants in the Informed group in comparison to the Uninformed group on the non-endpoint pairs, BD and BE, but was comparable between the groups on the non-endpoint pair, CE. Furthermore, and in contrast to Experiment 2, participants in the Uninformed group performed above chance levels on the non-endpoint pairs, BD and BE.

Statistical Analyses

A McNemar test revealed no significant difference in accuracy for the Informed group between the baseline and endpoint (p = .219) and non-endpoint pairs (p = .125). Similarly, no significant differences were observed in accuracy between the endpoint and non-endpoint pairs (p = 1.000) for participants in the Informed group.

A McNemar test revealed no significant differences in accuracy for the Uninformed group on the baseline and endpoint (p = .687), and non-endpoint pairs (p = .125). Similarly, no significant differences were observed in accuracy between the endpoint and non-endpoint pairs (p = .250) for participants in the Uninformed group.

Correlations between accuracy at test and post-experimental measures of awareness

In addition to comparing accuracy scores at test for participants in the Informed and Uninformed groups that did and did not meet test criterion, post-experimental measures of awareness were also correlated with test performances on all baseline, endpoint and non-endpoint pairs. In order to determine an awareness score for each participant, raters' were provided with a scoring guide for each question in the post-experimental awareness questionnaire. Each question was to be assigned a score out of a total of 2. Thus, raters were required to award a score of 2 for "definite awareness", on a specific question. A score of 1 was awarded for "some awareness", and a score of 0 was given for "no awareness". For instance, a score of 2 was awarded if a participant correctly ordered all five stimuli in the hierarchy, a score of 1 was awarded for two or less errors on this ordering, and a score of 0 was awarded if a participant made three or more errors on the ordering of the stimulus hierarchy. Each participant received a score out of 20 for their level of awareness. In order to calculate inter-observer agreement, the number of agreements between both raters was divided by the number of agreements + disagreements X 100. This calculation revealed that inter-observer agreement between both raters was high (83%). Pearson's correlations were conducted separately for participants in the Informed and Uninformed groups.

For the Informed group, Pearson's correlations revealed a positive correlation between task awareness and the baseline pair AB $r(16) = .022, p < .05$. Awareness was not correlated with accurate test performances on any other test pair (see Table 2.7). For the Uninformed group, Pearson's correlations revealed a positive correlation between task awareness and the baseline pair AB $r(13) = .031, p < .05$, and the endpoint pair AF $r(13) = .022, p < .05$. No other correlations between test performances and awareness of the stimulus hierarchy were noted (see Table 2.7).

In summary, providing participants in the Informed group with additional instructions at the start of the experiment, did not correlate with accurate performances at test.

Table 2.7
Pearson's correlations for awareness scores and test performances for participants in the Informed and Uninformed groups, that did and did not meet test criterion.

Pair	Informed	Uninformed
Baseline pairs		
AB	.510*	.482*
BC	.243	-.243
CD	.096	.174
DE	-.037	-.394
EF	-.305	.238
Endpoint pairs		
AC	.001	.150
AD	-.47	.057
AE	.067	.139
AF	-.215	.508*
BF	.051	.132
CF	-.300	.095
DF	-.133	.070
Non-endpoint pairs		
BD	.205	-.140
BE	-.092	.120
CE	-.144	-.052

* $p < .05$

In summary, the results of Experiment 3 demonstrated that sixteen participants from the Informed group and thirteen from the Uninformed group, successfully met criterion at testing. In addition, four participants from the Informed group and three from the Uninformed group failed to meet criterion at testing. High accuracy was maintained for both groups on the baseline pairs, with accuracy somewhat lower on the endpoint, and non-endpoint pairs. For example, similar to Experiment 2, high accuracy was observed for the Informed group on all test pairs. However, in comparison to Experiment 2, improvements in accuracy were noted for participants in the Uninformed group on the non-endpoint pairs. For example, and in contrast to Experiment 2, participants in the Uninformed group performed above chance on all of the non-endpoint pairs. Accuracy on the non-endpoint pairs BD and BE for participants in the Uninformed group in Experiment 2 was 54% and 63%, respectively. In contrast, accuracy for the Uninformed group on these same test pairs in Experiment 3 was 64% (BD) and 72% (BE). In addition, higher accuracy was observed for the Uninformed group on the non-endpoint pair, CE (91%) in Experiment 3 over that observed in Experiment 2 (86%). One potential reason for the improvements noted in response accuracy for the Uninformed group in Experiment 3 may be due to procedural differences between Experiments 2 and 3. For example, in contrast to Experiment 2, participants in Experiment 3 were exposed to unreinforced probe trials for the non-endpoint, inferential pair, BD, during training. Thus, presenting inferential, non-endpoint probe trials throughout training blocks may have facilitated more accurate responding on other inferential trials at test. The General Discussion will further explore this issue.

Experiment 3 also examined correlations between performance accuracy at test and post-experimental measures of awareness separately, for the Informed and Uninformed groups. For the Informed group, awareness of the stimulus hierarchy was only correlated with successful test performances on the AB baseline pair. No other correlations between task awareness and performance accuracy were noted. Thus, accurate test performances on the BD, BE and CE inferential pairs were not correlated with task awareness for the Informed group. For the Uninformed group, awareness of the stimulus hierarchy was correlated with successful test performances on the AB baseline pair and the AF endpoint pair. Performances on the BD, BE and CE inferential test pairs, were not correlated with task awareness. Thus, findings from the Informed group in Experiment 3, suggest that providing participants with additional instructions at the start of the experiment, does not guarantee awareness at the end of the experiment.

It must also be noted that, similar to findings from Experiments 1 and 2, for participants that failed to meet test criterion (Informed: P10, P15, P18 and P19; Uninformed: P5, P6 and P10), accuracy on the endpoint and non-endpoint pairs was below, or just above chance levels (Informed: P10, P15 and P18; Uninformed: P6, P8 and P10). More specifically, for a number of these participants, accuracy on the non-endpoint pairs ranged between 0% and 33% (Informed: P15, P18 and P19; Uninformed: P5, P6 and P10). Thus, despite the implementation of a number of interventions aimed at generating more accurate responding at testing (e.g., repeated exposure to training and test phases, and BD probe trials throughout training blocks), accuracy was still below criterion performance for a number of participants. The General Discussion will explore this issue in more detail.

General Discussion

The current chapter was concerned with examining the emergence of TI in adult participants. Experiments 1-3 considered some of the methodological features that may influence inferential responding. For example, Experiments 1-3 examined the potential utility of adopting test mastery criterion and repeated exposure to training and test phases on the emergence of TI. In addition, Experiments 2 and 3 considered the role of awareness, by examining performance accuracy on endpoint and non-endpoint pairs for participants that were provided with additional instructions at the start of the experiment (Informed), against a group that were not (Uninformed).

Results from Experiment 1 demonstrated that thirteen out of twenty-nine participants passed the experimental task, with only four participants displaying high levels of accuracy on the baseline, endpoint and non-endpoint pairs (P1, P19, P28 and P29). The remaining participants that passed, displayed high levels of accuracy on the baseline and endpoint pairs, but not on the non-endpoint pairs (P8, P9, P10, P11, P13, P14, P15, P17 and P20) In addition, a number of participants failed to meet criterion following additional exposure to training and test phases. For participants that failed to meet test criterion, high accuracy was maintained on the baseline pairs, but was somewhat lower on the endpoint (32%-71%) and non-endpoint pairs (25%-42%).

Experiment 2 sought to determine whether more accurate responses could be achieved by examining the role of awareness on inferential performances at test. More specifically, Experiment 2 sought to determine whether providing participants with additional instructions at the start of the study, would result in improvements in accuracy scores over those noted in Experiment 1. Performance accuracy was compared for two groups of participants (Informed and Uninformed), and findings revealed that high accuracy was observed on the baseline, endpoint and non-endpoint pairs for the Informed group. In comparison, high accuracy was observed for the Uninformed group on the baseline pairs, but was somewhat lower on the endpoint and non-endpoint pairs. More specifically, accuracy for the Uninformed group on the critical inferential (non-endpoint) pair BD was only just above chance levels (54%), while accuracy on the BE endpoint pair was slightly higher (63%). In contrast, accuracy for the Informed group on these same test pairs was high (BD: 81%; BE: 76%) However, high accuracy was observed for both the Informed and Uninformed groups on the non-endpoint pair CE (Informed: 91% Uninformed: 86%). Thus, findings from Experiment 2 suggest that providing participants with additional instructions at the start of the experiment facilitates more accurate responding at test. However, due to experimenter error with the post-experimental questionnaires, it was not possible to determine the extent to which pre-experimental instructions exerted an influence over inferential performances at test, and thus, Experiment 3 sought to explore this issue.

Experiment 3 also sought to compare performance accuracy at test for an Informed and Uninformed group of participants. However, in comparison to Experiment 2, additional measures of awareness were taken by means of post-experimental questionnaires. Findings revealed that high accuracy was again observed for both groups on the baseline pairs, with the Informed group displaying more accurate performances than the Uninformed group on the endpoint and non-endpoint pairs. However, improvements in performance accuracy were noted for the Uninformed group on the critical inferential (non-endpoint) pairs BD (64%) and BE (71%), in comparison to Experiment 2 (BD: 54%; BE; 63%). Accuracy for the Informed group in Experiment 3 on these same test pairs was however, higher (BD 72.5%; BE: 77.5%). Thus, similar to findings from Experiment 2, results from Experiment 3 suggest that providing participants with additional instructions at the start of the experiment, leads to more accurate responding at test. Experiment 3 also sought to further determine the influence of providing participants with additional instructions, by examining correlations between performance accuracy and responses to post-experimental questionnaires for participants in the Informed and Uninformed groups. Findings revealed that post-experimental measures of awareness were not correlated with responding on any of the inferential (non-endpoint) pairs for participants in the Informed and Uninformed groups. Findings from Experiment 3 therefore appear to support those of Lazareva and Wasserman (2010), who found that providing participants with information regarding the underlying stimulus hierarchy at the start of the experiment, does not guarantee awareness at the end of the study.

Test mastery criterion

A primary aim of Experiments 1-3 was to examine response accuracy for adult participants to novel inferential trials in a 6-term series. Across all experiments, a simultaneous discrimination was employed to examine the emergence of TI, but with some important procedural differences to the method in which TI is currently studied. For example, in Experiments 1-3, a mastery criterion of a minimum mean of 80% correct was employed during testing, which, if not met, resulted in re-exposure to training and test phases. Results demonstrated that, of the twenty-nine participants that started Experiment 1, thirteen met criterion at testing. In addition, eight participants were unable to complete the experiment, and a further eight participants failed to meet criterion at testing. Furthermore, in Experiment 2, sixteen out of twenty participants in the Informed group met criterion at testing, while twelve out of twenty participants in the Uninformed met criterion at testing. Results from Experiment 3 were similar to Experiment 2, in that sixteen out of twenty participants in the Informed group met criterion at testing, while thirteen out of twenty participants in the Uninformed group met test criterion.

As mentioned, currently studies examining TI do not employ mastery criterion during test phases. However, a common feature of behavioural studies is to incorporate mastery criterion during test phases, to determine whether stable patterns of responding have been established during training (see Dymond & Rehfeldt, 2000). For instance, previous studies have reported that inferential responding is said to have emerged if participants achieve an accuracy score ranging between 80% and 100% (e.g., Frank et al., 2005; Greene et al., 2001, 2006; Lazareva & Wasserman, 2010; Moses et al., 2006, 2008). However, if a specified accuracy mastery criterion was employed during test phases, then researchers may be provided with a more reliable method of determining the emergence of inferential responding. Furthermore, with the mastery criterion employed in the current study, participants were required to achieve a minimum mean of 80% correct across all test pairs. That is, across the baseline, endpoint, and non-endpoint pairs, participants were required to achieve an average of 80% (i.e., 48 out of 60 correct responses) across fifteen test pairs. However, a potential problem associated with incorporating an average accuracy mastery criterion, is that, participants may demonstrate accurate responding on some test pairs, but fail to respond accurately to other pairs. Indeed, findings from a number of participants in Experiment 1 that met test criterion revealed that accuracy on the non-endpoint pairs was at, or below chance levels (e.g., P2, P4 and P14). Thus, the mastery criterion employed throughout Experiments 1-3 allowed a number of participants to be classified as having passed the experimental task, despite the fact that they failed to demonstrate accurate responding on the inferential test pairs. Therefore, it may be beneficial for future studies to incorporate an accuracy mastery criterion on each test pair. That is, if each test pair is presented four times each, then it could be proposed that in order to determine whether inferential responding has emerged, participants are required to make a minimum of three out of four (75%) correct responses on each test pair. In turn, this may provide a more reliable method of determining whether inferential responding has emerged, and future studies should seek to take this into consideration.

Repeated exposure to training and testing

A common purpose of all experiments was to explore the effectiveness of repeated exposure to training and testing in facilitating the emergence of inferential responding in adult participants. As mentioned earlier, currently in studies examining the emergence of TI, participants are exposed to only one test block, in which all test pairs are presented (e.g., Frank et al., 2005; Greene et al., 2001; Moses et al., 2006). In contrast, a common feature of behavioural studies is to adopt a pre-determined test mastery criterion, which allows the untrained performances to emerge within a pre-determined number of exposures to testing. In effect, adopting this criterion provides researchers with the opportunity to examine the conditions necessary for the emergence of stable patterns of responding (e.g., Dymond & Rehfeldt, 2000). This in turn may have important implications when such problem-solving repertoires are found to be weak, in that repeated exposure to training and testing may allow the predicted patterns of behaviour to emerge over time.

With respect to the current studies, findings from Experiment 1 demonstrated that repeated exposure to training and test phases allowed a number of participants to meet test criterion, after having initially failed to do so. Furthermore, for a number of participants, accuracy on the endpoint and non-endpoint pairs, improved across additional exposures to training and test phases (e.g., P2, P4, P14, P19 and P28). Similarly, in Experiments 2 and 3, accuracy improved for a number of participants in the Informed and Uninformed groups on the endpoint and non-endpoint test pairs as a result of exposure to additional training and test phases (Experiment 2: Informed: P10, P13, P14, P18 and P19; Uninformed: P7, P9, P15, P16 and P19; Experiment 3: Informed: P5, P7, P12 and P17; Uninformed: P8, P11, P13, P15, P16, P17 and P18). Thus, findings from Experiments 1-3 revealed that, for a number of participants, repeated exposure to training and test phases allowed the predicted patterns of performance to emerge gradually, and over time. However, in order to more fully determine the potential utility of exposure to additional training and test phases, it may be necessary for future studies to compare initial test performances for participants that required additional exposure to training and test phases, to the exposure to testing that they met criterion. This in turn may help to more clearly determine the potential facilitative effects associated with this method of training and testing.

Furthermore, if, as was noted for a number of participants across Experiments 1-3, the predicted patterns of behaviour failed to emerge following the pre-determined number of exposures to training and testing, then it may be beneficial for researchers to develop and incorporate appropriate interventions to remediate these weaknesses. For example, and similar to Experiment 3 in the current study, it may be advantageous to expose these participants to unreinforced probe trials involving inferential pairs throughout the course of training. However, in the current study, probe trials were employed for all participants irrespective of whether the predicted patterns of responding did or did not emerge. Therefore, future research should seek to examine the effectiveness of presenting inferential probe trials throughout training, only for participants that are weak in their inferential repertoires.

Effects of prior instructions on inferential responding

As mentioned, Experiments 2 and 3 were concerned with the effects of pre-experimental instructions (i.e., prior awareness) on performance accuracy to baseline, endpoint and non-endpoint pairs at test. In Experiments 2 and 3, awareness was examined by comparing performance accuracy for a group of participants that received additional instructions at the start of the experiment that the stimuli could be arranged in a hierarchy (Informed), against a second group that were not given these instructions (Uninformed). Findings from Experiment 2 revealed that the Informed group displayed higher levels of accuracy on the endpoint and non-endpoint pairs, in comparison to the Uninformed group. In Experiment 3, findings again revealed that higher accuracy was observed for participants in the Informed group on the endpoint and non-endpoint test pairs, but that, in comparison to Experiment 2, participants in the Uninformed group performed above chance levels on the critical non-endpoint pairs, BD (64%) and BE (71%). Thus, taken together, findings from Experiments 2 and 3 suggest that awareness of the stimulus hierarchy has a facilitative effect on inferential performances at test.

The finding that response accuracy improved for participants in the Uninformed group in Experiment 3 warrants discussion. For example, procedural differences between Experiments 2 and 3 may have accounted for these findings. In contrast to Experiment 2, four unreinforced probe trials for the non-endpoint pair, BD, were presented during training blocks in Experiment 3. The incorporation of probe trials during training blocks is based on a previous study by Greene et al. (2001), who sought to determine whether presenting BD probe trials at the end of training blocks, would allow us to more clearly determine the relationship between inferential performances at test, and awareness of the task. In contrast, in Experiment 3 of the current thesis, unreinforced BD probe trials were presented throughout training blocks in an attempt to determine whether this would facilitate responding to other inferential trials (e.g., BE and CE) at test. The noted improvements in performance accuracy for participants in the Uninformed group on the BD and BE non-endpoint pairs, appear to suggest that the incorporation of BD probe trials throughout training blocks may have a facilitative effect on performances. However, in order to more fully determine the potential facilitative effects associated with this method of testing, it may be necessary for future studies to undertake a comparison of performance accuracy on the critical non-endpoint pairs for participants in the Uninformed group that were and were not exposed to these probe trials during training phases.

In addition, the finding from Experiments 2 and 3 that awareness of the stimulus hierarchy resulted in more accurate performances on novel endpoint and non-endpoint test pairs for participants in the Informed group warrants further discussion. For example, considerable debate exists among the literature on TI as to whether awareness is necessary for individuals to respond to inferential problems at test. For instance, Frank et al. (2005) and Greene et al. (2001) found evidence for the expression of TI in adult humans in the absence of explicit awareness (see also Frank et al., 2005), whereas Moses et al. (2006) and Lazareva and Wasserman (2010) report that awareness of the stimulus hierarchy is necessary for successful inferential responding. Furthermore, findings from the literature on TI suggest that if accuracy is higher on the test pairs containing end terms (e.g., AC and BF) over those that do not contain end terms (e.g., BD and BE) then such findings are indicative that participants employ lower-level associative learning strategies to solve the TI task (e.g., Bryant & Trabasso, 1971; von Fersen et al., 1991; Wynne, 1995, 1997). With respect to findings from Experiments 1-3 in the current chapter, results from Experiment 1 revealed that participants demonstrated patterns of responding that were indicative of associative-learning strategies. That is, accuracy was significantly higher on the endpoint test pairs in comparison to the non-endpoint test pairs. In contrast, such patterns of performance were not observed for participants in the Informed and Uninformed groups in Experiments 2 and 3. Thus, it is difficult to determine why participants in Experiment 1 displayed patterns of performance indicative of lower-level associative learning strategies, while participants in the Uninformed groups in Experiments 2 and 3, did not, when similar training and testing protocols were employed. Further research is therefore warranted on this issue, as findings from participants in the current thesis provide conflicting evidence regarding the strategies that humans employ to solve the TI task.

With respect to the method in which awareness was assessed in the current thesis, participants were assigned to either an Informed or Uninformed group at the start of the experiment, similar to the Greene et al. (2001) study. The Informed group received additional instructions at the start of the experiment that the stimuli could be arranged in a hierarchy, while the Uninformed group did not receive these instructions. Greene et al. (2001) reported that the Uninformed group were capable of responding to the BD pair in the absence of explicit awareness. However, Greene et al. (2001) did report that accuracy was higher on the BD probe trial for participants in the Informed group (Experiment 1: 98%), in comparison to those in the Uninformed group (Experiment 1: 87%; see also Frank et al., 2005; Lazareva & Wasserman, 2010). With respect to the current findings, although participants in the Uninformed group in Experiment 3 were capable of responding above chance on the non-endpoint pairs, accuracy was not at the high levels reported by Greene et al. (2001). However, in the Greene et al. (2001) study, participants were exposed to training and testing with a 5-term series, which involved the presentation of only one inferential probe trial (BD) at testing. In contrast, in the current study, participants were exposed to training and testing with a 6-term series, and were exposed to a greater number of inferential trials at test (e.g., BE and CE). Thus, it is difficult to determine whether in the Greene et al. (2001) study, comparably high levels of performance accuracy would be observed for the Uninformed group if additional non-endpoint pairs were presented at test.

The current findings are however, similar to those reported by Moses et al. (2006, 2008), and Lazareva and Wasserman (2010). For example, Moses et al. (2006) reported that successful inferential performances at test were associated with the ability to report the underlying stimulus hierarchy. Furthermore, and similar to the findings from Experiment 3 of the current thesis, Lazareva and Wasserman (2010) reported that that some participants in their study were capable of responding to transitive tests in the absence of explicit awareness (Uninformed), but that awareness improves transitive responding (Informed). In addition, Lazareva and Wasserman (2010) found evidence that providing participants with additional instructions at the start of the experiment, does not guarantee awareness at the end of the experiment. That is, the authors found no significant difference in awareness between the Informed and Uninformed groups at the end of the experiment. Similarly, findings from Experiment 3 of the current study did not find that accurate responding on the inferential, non-endpoint pairs BD, BE and CE, were correlated with post-experimental measures for the Informed group. In turn, such findings seem important considering the debate regarding the role of awareness on the emergence of TI. Indeed, the disparity observed across studies regarding the role of awareness on TI, illustrates that perhaps alternative methods of examining the role of awareness on the emergence of TI are needed. For example, as mentioned in Chapter 1, the role of awareness may be more clearly identified by incorporating concurrent self-report measures during the task. This is comparable to behaviour-analytic research examining the role of private verbal behaviour on operant performances in adult humans (known as the "silent dog" protocol; Cabello, Luciano, Gomez, & Barnes-Holmes, 2004; Hayes, Brownstein, Haas, & Greenway, 1986; Rosenfarb, Newland, Brannon, & Howey, 1992). More specifically, this procedure seeks to determine whether verbal behaviour affects participants' ability to contact the programmed contingencies during schedule tasks (e.g., Cabello et al., 2004). For instance, Cabello et al. (2004) took concurrent measures of participants' self-reports whilst they were exposed to one of two schedules of reinforcement. Findings revealed that there was a significant correlation between specific types of self-reports (counting and describing) and performances on these schedules. That is, counting aloud facilitated participant responding in accordance with the relevant schedule. Thus, although these findings are only correlational, the authors propose that the protocol has the potential to identify the factors governing human behaviour on operant tasks. Therefore, it may be beneficial for future studies seeking to examine the role of awareness on the emergence of TI, to incorporate such measures.

In conclusion, findings from the current chapter highlight the potential utility of incorporating mastery criterion during test phases, in studies examining TI. In addition, the current findings also highlight the potential utility of exposing participants to additional training and test phases, if the predicted patterns of performance do not emerge immediately. Thirdly, findings from Experiments 2 and 3 suggest that awareness, leads to more accurate responding at test. However, findings from Experiment 3 revealed that providing participants with additional instructions at the start of the experiment does not guarantee awareness at the end, and thus, it may be necessary for researchers to develop and incorporate alternative methods to determine the specific role of awareness on the TI task.

Chapter 3

Developing a Novel Behaviour-Analytic Account of Transitive Inference with the Relational Frame of Comparison

The current chapter sought to explore the utility of a novel account of TI based on the principles of Relational Frame Theory. Indeed, findings from research on derived comparative relations compare favourably with findings from the literature on TI (e.g., Dymond & Barnes, 1995; Hinton et al., 2010; Munnelly et al., 2010; O'Hora et al., 2002; Reilly et al., 2005; Whelan et al., 2006; see Chapter 1). With this approach, verbally-able humans are trained on a number of overlapping conditional discriminations, such as B MORE-THAN A, and C MORE-THAN B. Later, during testing, participants may derive that A is LESS-THAN B and B is LESS-THAN C (mutual entailment), and also that C is MORE-THAN A, and A is LESS-THAN C (combinatorial entailment), in the absence of further training. Thus, similar to research on TI, untrained relations typically emerge between non-adjacent stimuli, following training on adjacent stimulus pairings.

Current Experiments

The current chapter sought to replicate and extend previous findings from Munnelly et al. (2010) and Reilly et al. (2005), who employed the conditional discrimination outlined above, to examine derived comparative responding. Both studies found that the procedure could successfully establish derived comparative responding in adult humans, and propose that such patterns of responding could potentially be employed as a novel account of TI. In addition, both studies found that the account has the ability to examine some of the characteristic effects (e.g., SDE; Reilly et al., 2005; see also O'Hora et al., 2002), and factors (e.g., linearity; Munnelly et al., 2010), associated with the emergence of TI. However, further research is needed to explore the conditions under which this behaviour emerges. For example, Reilly et al. (2005) undertook an analysis of reaction times to different combinations of "More-than" and "Less-than" relations at test, whereas Munnelly et al. (2010) examined response accuracy to these relations. However, an issue arising from the Munnelly et al. study was that accuracy was low on the mutually entailed relations, and one- and two-node relations that were *different to training* in comparison to those that were the *same as training*. That is, participants who were trained on, for example, "More-than" relations and tested on a combination of "More-than" and "Less-than" relations, displayed lower levels of accuracy on the "Less-than" relations at test than on the "More-than" relations.

One potential reason for these findings may centre on the fact that in the Munnelly et al. (2010) study, participants were only exposed to one presentation of each test pair. For example, individual participants may have made either a correct or incorrect response to each stimulus, leading to response accuracy of 100% or 0%, respectively. When performance accuracy was then averaged for group analysis, this may have resulted in the average group accuracy falling to near chance levels (50%). Indeed, other studies examining the emergence of arbitrary comparative relations typically expose participants to a greater number of presentations of each test trial (e.g., 2; Whelan et al., 2006). Similarly, studies examining TI in adults using the simultaneous discrimination, have presented participants with between six and eight trials of each test problem (e.g., Frank et al., 2005; Greene et al., 2001; Lazareva & Wasserman, 2010; Moses et al., 2006). Thus, increasing the number of trials presented at test provides participants with a greater opportunity to achieve the pre-determined test mastery criterion. In turn, and as mentioned, presenting a larger number of test trials may allow us to determine whether participants are randomly making responses at test. That is, if participants select the correct stimulus on both occasions during testing (100%), then we may propose that effective stimulus control has been established over responding. However, if participants select the incorrect stimulus on both occasions, then such findings may reveal that stimulus control over responding is lacking. Experiment 4 sought to increase the number of presentations of each test trial in order to control for the potential confounds associated with chance performances.

Experiment 4 also sought to determine whether exposing participants to differing arbitrary relational training groups, impacts arbitrary comparative performances at test. For instance, at the start of the experiment, participants were randomly assigned to an All-More or All-Less training group, in which the arbitrary training pairs differed between the groups (All-More: B>A, C>B, D>C, E>D; All-Less: A<B, B<C, C<D, D<E). A previous study by Reilly et al. (2005) examined reaction times to a combination of "More-than" and "Less-than" problems at test, for three different training groups. One group of participants were exposed to arbitrary relational training with only "More-than" relations, while a second group received training with only "Less-than" relations and a third group received training with a commination of "More-than" and "Less-than" relations. All three groups were then

exposed to the same problems at test. Reilly et al. found that reaction times for the "All-More" training group were significantly faster on all test problems relative to the All-Less and Less-More training groups. The authors reported that the observed performances may be due in part to the proposal that "More-than" relations appear earlier than "Less-than" relations in our behavioural repertoires (see also Barnes-Holmes et al., 2004). However, Munnelly et al. (2010) found no differences in terms of response accuracy between the All-More and All-Less groups on any relational problems at test. Experiment 4 of the Chapter 3 therefore sought to undertake a similar analysis of the differing training groups.

Lastly, the current chapter adopted a conditional discrimination to examine the emergence of TI. For example, in Chapter 2, participants were exposed to a simultaneous discrimination in which both the reinforced and non-reinforced discriminative stimuli were presented simultaneously onscreen. The conditional discrimination paradigm employed in the current chapter is similar, with the exception that the function of the discriminative stimulus changes on the basis of the contextual cue presented. Thus, in Experiment 4, participants were first exposed to non-arbitrary relational training and testing to establish the contextual functions of MORE-THAN and LESS-THAN for two arbitrary images. Next, participants were exposed to arbitrary relational training, followed by arbitrary relational testing, which involved the presentation of a combination of "More-than" and "Less-than" test problems. Furthermore, and similar to Chapter 2, participants in Experiment 4 were exposed to additional training and testing phases if they initially failed to meet accuracy criterion during arbitrary relational testing.

Experiments 5A and 5B were also concerned with an examination of arbitrarily applicable comparative responding.

Experiment 4

Method

Participants

Twenty-three participants, four male and nineteen female, ranging in age from 18 to 22 years (M_{age} = 20.00, SD = 1.04) were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the study, and were randomly assigned to the All-More or All-Less training group at the start of the Experiment. Ethical approval was obtained from the Psychology Department Ethics Committee before research commenced.

Apparatus and Setting

The experiment took place in a research laboratory in the Department of Psychology at Swansea University. Participants were seated at a table in an experimental room (2 X 3 metres) containing a personal computer with a 16-inch display screen on which all training and testing trials were presented. All instructions were presented in white on a black background throughout the course of the experiment. The experiment was programmed using Presentation (Neurobehavioural Systems, Palo Alto, CA), which controlled stimulus presentations, and recorded all responses.

Materials and Stimuli

Two arbitrary visual stimuli were employed as contextual cues during non-arbitrary relational training and testing to establish the contextual functions of MORE-THAN and LESS-THAN (see Figure 3.1). In addition, eight non-arbitrary stimulus sets were employed and were composed of images of different quantities of particular objects.

Five three-letter, consonant-vowel-consonant nonsense syllables were employed (VEK, JOM, BIH, CUG and PAF) during arbitrary relational training and testing as comparison stimuli. From these arbitrary stimuli, a 5-member linear relational network was constructed which is described as follows: A-B-C-D-E (Note. Participants were not exposed to these labels, which are used here in the interests of clarity).

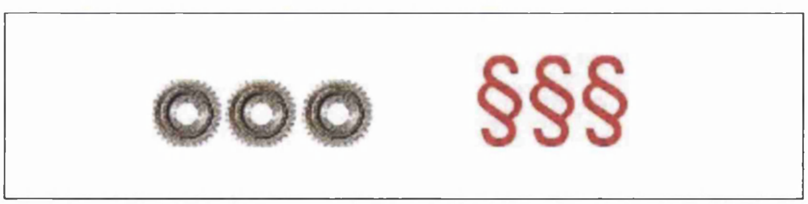


Figure 3.1. The two arbitrary stimuli employed as MORE-THAN and LESS-THAN contextual cues.

Procedure

The procedure consisted of two different training and testing phases, and is based on those employed by Munnelly et al. (2010): *Phase 1A: Non-arbitrary Relational Training* and *Phase 1B: Non-arbitrary Relational Testing* and *Phase 2A: Arbitrary Relational Training* and *Phase 2B: Arbitrary Relational Testing.*

For all phases, at the start of each trial, the contextual cue appeared in the centre top third of the computer screen. Following a delay of 1.5 s, two comparison stimuli appeared simultaneously in the lower third of the left- and right-hand side of the screen. The screen position (i.e., left or right) of these comparisons was counterbalanced across trials. Participant selections were made by pressing either the "z" or "/" (for the comparison on the left- or right-hand side of the computer screen, respectively). If a participant made a correct selection, the screen cleared and the word "Correct!" appeared in the middle of the computer screen. If a participant made an incorrect selection, the word "Wrong" appeared in the middle of the computer screen. Feedback was provided only for non-arbitrary and arbitrary relational training phases and was omitted for both test phases. Feedback was displayed for 1.5 s, and an inter-trial interval (ITI) of 1.5 s followed each trial. Both the contextual cue and the comparison stimuli remained onscreen until a response was recorded.

Phase 1A: Non-arbitrary Relational Training. The purpose of this phase was to establish contextual control for the two arbitrary cues (e.g., MORE-THAN and LESS-THAN) over participant responding to stimulus sets that varied in terms of their physical quantities. There were four stimulus sets employed during non-arbitrary relational training, which were composed of images of different quantities of particular objects. The quantities of objects were termed *Few* for the smallest amount, *Intermediate* amount (Note: not necessarily the midpoint of the smallest and greatest amounts) and *Many* for the greatest amount. For example, one stimulus set was composed of images of one, two and eight basketballs. The four stimulus sets employed were as follows (the quantities of the particular object that composed each image in parentheses): basketballs (1, 2, 8), beakers (1, 3, 6), tractors (1, 2, 3), and ladybirds (2, 4, 8). Each set was composed of three images and two contextual cues, generating the following six discriminations: LESS-THAN [Few/Intermediate], LESS-THAN [Few/Many], LESS-THAN [Intermediate/Many], MORE-THAN [Few/Intermediate], MORE-THAN [Many/Intermediate], and MORE-THAN

[Many/Few]. A total of twenty-four trial types were generated from these stimulus sets. Phase 1 began with the following instructions onscreen:

> During this phase you will be presented with one cue in the middle of the screen and two images beneath it in the centre of the screen, one on the right and one on the left. Your task is to choose one of the images. To select the image on the right, press the marked key on the right of the keyboard. To select the image on the left, press the marked key on the left of the keyboard. Please try to do so as quickly and as accurately as possible. Sometimes the computer will give you feedback, and at other times it will not. However, you can get all of the tasks without feedback correct by carefully attending to the tasks with feedback. Remember, there is always a correct answer. The computer will tell you when this phase is finished. Please press the space bar to begin.

As described above, the contextual cue appeared first in the centre top-third of the computer screen, with the two comparison stimuli appearing simultaneously following a 1.5 s delay. When the contextual cue for MORE-THAN was presented on the computer screen, choosing the comparison stimulus with the greater quantity produced the feedback "Correct!", while choosing the comparison with the lesser quantity, produced the feedback "Wrong". Similarly, when the contextual cue for LESS-THAN was presented, choosing the comparison with the lesser quantity produced the feedback "Correct!", and selecting the comparison with the greater quantity produced the feedback "Wrong". Non-arbitrary relational training continued until participants emitted 10 consecutive correct responses (see Figure 3.2).

Once the mastery criterion was met, participants immediately proceeded to the non-arbitrary relational test.

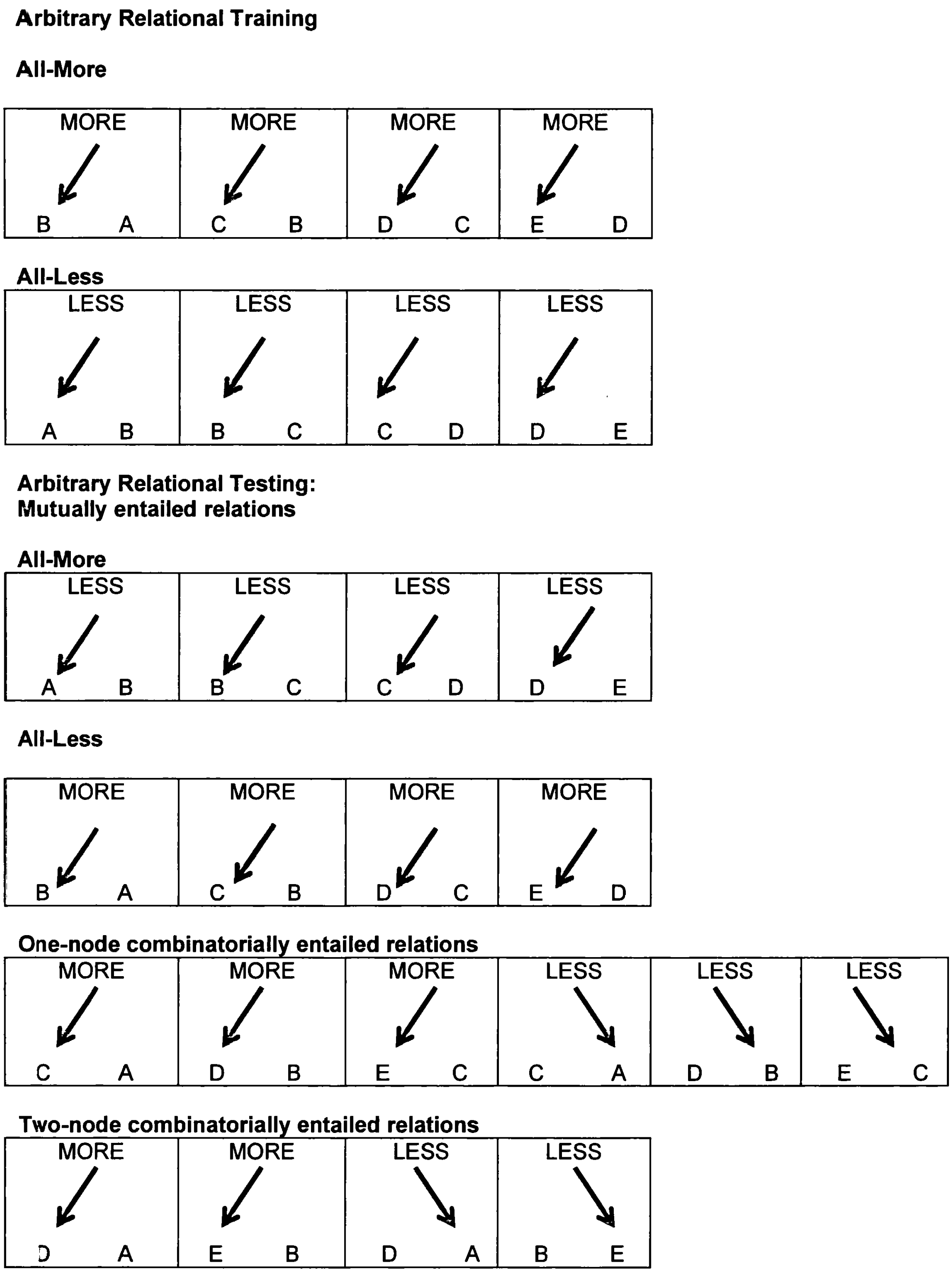


Figure 3.2. The arbitrary relational training and test trials that participants in the All-More and All-Less training groups were exposed to in Experiment 4. The red arrow points to the predicted correct response.

Phase 1B: Non-arbitrary Relational Testing. This phase was identical to Phase 1A with the exception that feedback was omitted. In addition, the test trials consisted of four novel stimulus sets, which were as follows: turtles (2, 3, 4), arks (1, 2, 3), apples (1, 4, 8), and traffic lights (1, 3, 4). Again, twenty-four trial types were generated from these stimulus sets. Mastery criterion for this phase was again set at 10 consecutive correct responses. If, however, this criterion was not met following exposure to twenty-four test trials, participants were re-exposed to non-arbitrary relational training, followed again by non-arbitrary relational testing.

Phase 2A: Arbitrary Relational Training. The following instructions appeared immediately onscreen following the completion of Phase 1 and signalled the beginning of Phase 2:

> The first phase of the experiment is finished. Thank you. During this phase you will be presented with one cue in the middle of the screen and two images beneath it in the centre of the screen, one on the right and one on the left. As before, your task is to choose one of the images. To select the image on the right, press the marked key on the right of the keyboard. To select the image on the left, press the marked key on the left of the keyboard. Please try to do so as quickly and as accurately as possible. Later, in the tasks without feedback, you will be presented with the cues that you have seen before. Please look at the cue, as you can use it to help you learn which one of the images below is the correct one to choose. You can get all of the tasks without feedback correct by carefully attending to the tasks with feedback. Remember, there is always a correct answer! The computer will tell you when you are finished. Please press the space bar to begin.

Similar to Phase 1, the contextual cue again appeared onscreen and was followed by the two comparison stimuli. However, in this phase, the comparison stimuli consisted of nonsense syllables, which are labelled A, B, C, D and E for ease of clarity. During this training phase, participants were presented with four training trials, and the training pairs differed between the All-More and All-Less groups. For example, the training pairs presented to the All-More group in the presence of the MORE-THAN contextual cue were: B>A, C>B, D>C and E>D (where ">" describes the contextual cue for the relation MORE-THAN. However, it is important to note that “>” and “<” are used here to denote the contextual cues of MORE-THAN and LESS-THAN, respectively. Participants were not exposed to these inequality symbols, but instead, two abstract visual images as contextual cues). This training resulted in the following relational network: E>D>C>B>A (see Figure 3.2 and Table 3.1). Similarly, for the All-Less group, the training pairs presented in the presence of the LESS-THAN contextual cue were: A<B, B<C, C<D and D<E (where “<” describes the contextual cue for the reinforced relation LESS-THAN), designed to result in the following relational network: A<B<C<D<E (see Table 3.1).

Each of the four training trials were presented in a quasi-random order, three times each, within a block of 12 trials. Participants were required to make 12 out of 12 (i.e., 100% accuracy) correct responses on a given training block to achieve training mastery criterion. Training blocks were repeated until this criterion was achieved.

Table 3.1
Training and Test Trials Received by the Two Groups in Experiment 4.

Group	Relation Type	Test Trial Type					
		Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
	ME	B>A	C>B	D>C	E>D		
All-More	Baseline	B>A	C>B	D>C	E>D		
	ME	A<B	B<C	C<D	D<E		
		Relations Common to All Groups					
	CE1	C>A	D>B	E>C	A<C	B<D	C<E
	CE2	D>A	E>B	A<D	B<E		

Note. “Baseline” refers to test trials involving directly trained relations, and the acronym ME, CE1 and CE2 refer to test trials for mutually entailed and one- and two-node combinatorially entailed relations, respectively. The inequality symbols, < (LESS-THAN) and > (MORE-THAN), denote the contextual cue that was presented: This indicates which comparison should be “selected over” the other, with the reinforced comparison to the left, and the punished comparison to the right of the inequality symbol. It is important to note that the actual contextual cues used in the present study consisted of abstract visual images, and not the inequality symbols described here, which are used for the purposes of clarity.

Phase 2B: Arbitrary Relational Testing. Upon completion of the arbitrary relational training phase, participants were immediately exposed to the arbitrary relational test phase. All feedback was omitted and participants were presented with the four baseline relations along with 14 novel test trials (see Figure 3.2 and Table 3.1). All 18 test trials were presented in a pseudo-random order, twice each, within a block of 36 test trials. The novel test trials included mutually entailed and one- and two-node combinatorially entailed relations. For example, the mutually entailed test trials presented to the All-More group were: A<B, B<C, C<D and D<E, and for the All-Less group, the mutually entailed test trials were: B>A, C>B, D>C and E>D. Both groups were presented with the following six one-node combinatorially entailed relations: C>A, D>B, E>C, A<C, B<D and C<E. Similarly, both groups received the same two-node combinatorially entailed relations (D>A, E>B, A<D and B<E). The endpoint pairing of A and E was not presented during this test, as A would always be preferred over E, as a result of their direct reinforcement history (Vasconcelos, 2008).

The pre-determined mastery criterion for the arbitrary relational test was a minimum of 30 correct responses out of a total of 36 test trials (i.e., 83% accuracy). If this mastery criterion was not achieved, then the following instructions appeared onscreen, and participants were re-exposed to both the non-arbitrary and arbitrary relational training and testing phases again, from the very beginning, for a maximum of three more times:

> Please take a break. You will now be re-exposed to the experimental tasks because your choices during the tasks without feedback did not meet criteria. Please pay special attention to everything onscreen and use what you learn during the choices with feedback to solve the choices without feedback. Please press the space bar to begin.

On the other hand, if a participant was successful in reaching this criterion, then this signalled the end of the experiment, where the following instructions appeared onscreen:

> You're done! Thank you for taking part.

Results and Discussion

Of the twenty-three participants that started Experiment 4, three were unable to progress beyond the non-arbitrary relational training phase, and their data is therefore excluded from further analyses. For the remaining twenty participants, four from the

All-More group and one from the All-Less group failed to achieve the pre-determined arbitrary relational test criterion within the maximum four test exposures. However, six participants from the All-More group and nine from the All-Less group met criterion on the arbitrary relational test, and required between 1 and 4 ($M = 2.33$, $SD = 1.11$; see Table 3.2 and Table 3.3) exposures to do so. Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Training trials to criterion

Table 3.2 and Table 3.3 present a summary of group (All-More and All-Less) performances during the non-arbitrary and arbitrary relational training and testing phases, for participants that passed and failed the experiment. Participants in the All-More training group required, on average, a greater number of training trials during the non-arbitrary relational training phase, compared to the All-Less training group However, this difference was non-significant ($t(13) = .88$, $p = .40$). An analysis of the number of training trials to achieve mastery criterion for the arbitrary relational training phase again revealed no significant differences between the groups ($t(13) = -.66$, $p = .52$). Participants in the All-More and All-Less groups that passed and failed the experiment were exposed to the experimental task, on average, 3.3 ($SD = 1.06$), and 2.10 ($SD = 1.98$) times, respectively.

Accuracy: Baseline relations

Six participants from the All-More training group (P1, P3, P11, P15, P17 and P19), and nine participants (P2, P4, P6, P8, P10, P12, P14, P16 and P20) from the All-Less group met criterion during arbitrary relational testing. A graphical representation of performances on the baseline relations for participants in the All-More and All-Less groups that passed and failed the experiment can be seen in Figure 3.3. This graph shows that high accuracy was observed on the baseline pairs for both groups, with participants in the All-Less group displaying slightly higher levels of accuracy that the All-More group on the baseline relations. Average accuracy for the All-More group on the baseline relations ranged between 90% and 95% and between 95% and 100% for the All-Less group.

Table 3.2.
Individual data for participants in the All-More group that passed and failed Experiment 4.

Participant	Phases 1: Non-arbitrary Relational Training	Phase 2: Arbitrary Relational Training	Phase 3: Arbitrary Relational Testing B	ME	CE1	CE2
1	301	120	7/8	7/8	12/12	8/8
3	197	48	8/8	0/8	6/12	4/8
	14	12	7/8	1/8	6/12	4/8
	15	12	8/8	0/8	6/12	4/8
	10	12	8/8	8/8	12/12	8/8
5*	55	24	7/8	5/8	4/12	6/8
	22	24	7/8	5/8	8/12	5/8
	24	48	7/8	7/8	3/12	4/8
	18	12	8/8	3/8	8/12	5/8
7*	21	120	8/8	7/8	6/12	3/8
	10	24	8/8	7/8	3/12	1/8
	10	48	8/8	7/8	6/12	4/8
	17	36	8/8	7/8	3/12	5/8
9*	12	96	7/8	0/8	5/12	5/8
	10	24	8/8	7/8	4/12	1/8
	10	12	8/8	8/8	3/12	1/8
	10	12	7/8	7/8	4/12	0/8
11	34	84	7/8	0/8	5/12	4/8
	10	24	8/8	0/8	10/12	3/8
	10	24	8/8	0/8	7/12	4/8
	10	48	8/8	8/8	12/12	7/8
13*	15	36	5/8	3/8	4/12	7/8
	10	36	7/8	4/8	4/12	4/8
	14	24	6/8	5/8	5/12	3/8
	10	12	4/8	3/8	4/12	3/8
15	21	108	8/8	1/8	8/12	5/8
	10	48	8/8	0/8	6/12	4/8
	19	24	7/8	8/8	12/12	8/8
17	42	24	3/8	4/8	5/12	4/8
	10	12	5/8	4/8	4/12	4/8
	10	60	8/8	8/8	9/12	6/8
19	33	48	6/8	0/8	6/12	5/8
	10	48	8/8	7/8	11/12	8/8

Note. Data is shown for the number of trials required to achieve mastery criterion for the non-arbitrary relational training and testing phases, and also for the arbitrary relational training phase. Number of correct responses on each exposure to the baseline, mutually entailed, and one- and two-node combinatorially entailed relations is also shown. The acronym B, ME, CE1 and CE2 refer to test trials for baseline, mutually entailed and one- and two-node combinatorially entailed relations, respectively. * refers to participants that failed to meet criterion on the arbitrary relational test.

A summary of performance accuracy for participants in the All-More group across each exposure to testing can be seen in Table 3.2. When exposed to the baseline relations, four participants from the All-More group that met criterion at testing (P3, P11, P17 and P19), made no errors, while two participants made one error (P1 and P15). For the four participants in the All-More group that failed to meet test criterion (P5, P7, P9 and P13), two (P5 and P7) made no errors on the baseline relations, one (P9) made one error, and one participant (P13) made four errors. A summary of performance accuracy for participants in the All-Less group across each exposure to testing can be seen in Table 3.3.

When exposed to the baseline relations, seven participants in the All-Less group that met criterion at testing (P2, P4, P8, P10, P12, P14 and P16) made no errors, while two participants (P6 and P20) made two errors. For the one participant in the All-Less group (P18) that failed to meet test criterion, P18 made no errors on the baseline relations. In summary, irrespective of whether participants passed or failed the experiment, high accuracy was observed for the All-More and All-Less groups on the baseline relations.

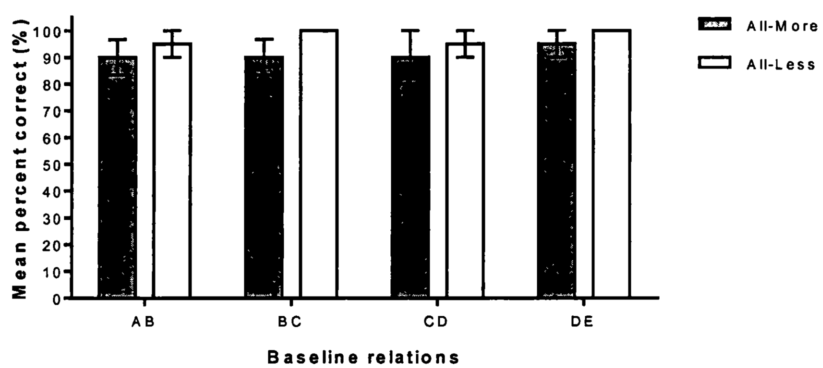


Figure 3.3. Experiment 4: The mean percent correct for participants in he All-More and All-Less groups to the baseline relations. Error bars indicate standard errors.

Table 3.3.
Individual data for participants in the All-Less group that passed and failed Experiment 4.

	Phases 1: Non-arbitrary Relational	Phase 2: Arbitrary Relational	Phase 3: Arbitrary Relational Testing			
Participant	Training	Training	B	ME	CE1	CE2
2	18	72	6/8	3/8	8/12	5/8
	23	36	8/8	0/8	7/12	4/8
	12	12	8/8	8/8	12/12	8/8
4	19	96	1/8	0/8	2/12	0/8
	10	72	8/8	8/8	11/12	8/8
6	21	132	7/8	7/8	12/12	6/8
8	22	168	6/8	2/8	7/12	5/8
	10	36	8/8	6/8	11/12	8/8
10	92	48	8/8	8/8	12/12	8/8
12	28	36	8/8	8/8	11/12	8/8
14	32	72	8/8	0/8	7/12	3/8
	10	48	8/8	7/8	9/12	7/8
16	38	60	8/8	8/8	10/12	8/8
18*	14	132	5/8	0/8	6/12	4/8
	10	12	6/8	2/8	6/12	4/8
	10	24	8/8	0/8	6/12	4/8
	20	12	8/8	0/8	6/12	4/8
20	14	84	1/8	0/8	0/12	1/8
	10	24	8/8	4/8	10/12	6/8
	10	36	6/8	6/8	10/12	7/8
	10	12	7/8	7/8	12/12	7/8

Note. Data is shown for the number of trials required to achieve mastery criterion for the non-arbitrary relational training and testing phases, and also for the arbitrary relational training phase. Number of correct responses on each exposure to the baseline, mutually entailed, and one- and two-node combinatorially entailed relations is also shown. The acronym B, ME, CE1 and CE2 refer to test trials for baseline, mutually entailed and one- and two-node combinatorially entailed relations respectively. * refers to participants that failed to meet criterion on the arbitrary relational test.

Mutually entailed relations. A graphical representation of performances on the mutually entailed relations for participants in the All-More and All-Less groups that passed and failed the experiment can be seen in Figure 3.4. This graph shows that high accuracy was observed on the mutually entailed relations, and was, for the most part, comparable between the groups. However, it must be noted that accuracy was higher for the All-More and All-Less groups on the AB and CD mutually entailed relations, in comparison to the BC and CD pairs. Average accuracy for the All-More group on the mutually entailed relations ranged between 70% and 90% and between 75% and 90% for the All-Less group.

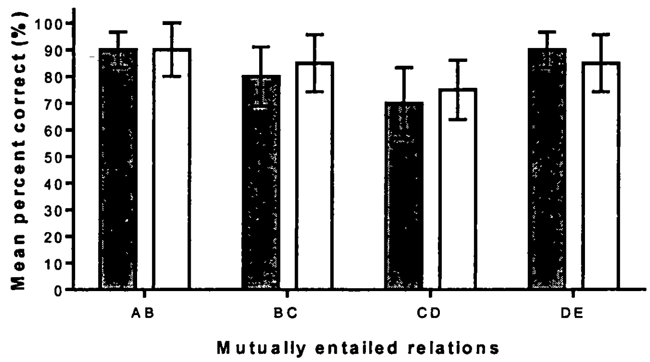


Figure 3.4. Experiment 4: The mean percent correct for participants in the All-More and All-Less groups to the mutually entailed relations. Error bars indicate standard errors.

A summary of performance accuracy for participants in the All-More group across each exposure to testing can be seen in Table 3.2. When exposed to the mutually entailed relations, four participants in the All-More group that met criterion during testing (P3, P11, P15 and P17), made no errors, while two participants made one error (P1 and P19). For the four participants (P5, P7, P9 and P13) in the All-More group that failed to meet test criterion, two participants (P7 and P9) made one error on the mutually entailed relations, and two participants (P5 and P13) made five errors.

A summary of performance accuracy for participants in the All-Less group across each exposure to testing can be seen in Table 3.3. When exposed to the mutually entailed relations, five participants in the All-Less group that met criterion at testing, made no errors (P2, P4, P10, P12 and P16), while three participants made one error (P6, P14 and P20), and one participant (P8) made two errors. For the one participant (P18) that failed to meet criterion during the arbitrary relational test, P18 made no correct responses on the mutually entailed relations. In, summary, high accuracy was observed for both the All-More and All-Less groups on the mutually entailed relations, with accuracy slightly higher for both groups on the AB and DE pairs, in comparison to the BC and CD pairs.

One-node relations. A graphical representation of performances on the one-node combinatorially entailed relations for participants in the All-More and All-Less groups that passed and failed the experiment can be seen in Figure 3.5. This graph shows that high accuracy was observed on the C>A, and A<C one-node relations for both groups. Slightly lower accuracy was observed for both groups on the B<D and C<E one-node relations. Furthermore, accuracy on the one-node pairs D>B and E>C was considerably higher for participants in the All-Less group in comparison to the All-More group. Average accuracy for the All-More group on the one-node relations ranged between 60% and 90% and between 65% and 93% for the All-Less group.

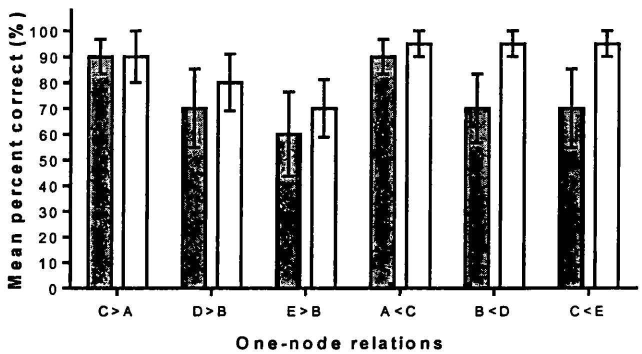


Figure 3.5. Experiment 4: The mean percent correct for participants in the All-More and All-Less groups to the one-node combinatorially entailed relations. Error bars indicate standard errors.

A summary of performance accuracy across each exposure to testing for participants in the All-More group that passed and failed the experiment can be seen in Table 3.2. With respect to the one-node relations, four participants (P1, P3, P11 and P15) from the All-More group that met criterion during arbitrary relational testing, made no errors, while one participant made one error (P19), and one participant made three errors (P17). For the four participants in the All-More group (P5, P7, P9 and P13) that failed to meet test criterion, one participant (P4) made four errors on the one-node relations, two participants (P9 and P13) made eight errors, and one participant (P7) made nine errors.

A summary of performance accuracy across each exposure to testing for participants in the All-Less group that passed and failed the experiment can be seen in Table 3.3. With respect to the one-node relations, four participants from the All-Less group that met criterion during the arbitrary relational test, made no errors (P2, P6, P10 and P20), three participants made one error (P4, P8 and P12), one participant made two errors (P16), and one participant made three errors (P14). For the one participant (P18) from the All-Less group that failed to meet test criterion, P18 made six errors on the one-node relations. In summary, accuracy on the one-node relations varied for both groups, but for the most part, the All-Less group, outperformed the All-More.

Two-node relations. A graphical representation of performances on the two-node combinatorially entailed relations for participants in the All-More and All-Less groups that passed and failed the experiment can be seen in Figure 3.6. This graph shows that high accuracy was observed for the All-Less group on the two-node relations, while accuracy was lower for the All-More group on some of these test pairs. Average accuracy for the All-More group on the two-node relations ranged between 65% and 80% and between 86% and 93% for the All-Less group.

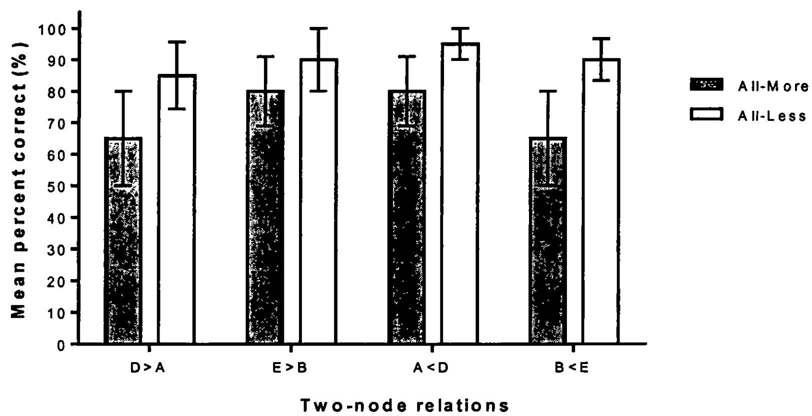


Figure 3.6. Experiment 4: The mean percent correct for participants in the All-More and All-Less groups to the two-node combinatorially entailed relations. Error bars indicate standard errors.

A summary of performance accuracy across each exposure to testing for participants in the All-More group that passed and failed the experiment can be seen in Table 3.2. With respect to the two-node relations, four participants from the All-More group that met criterion during the arbitrary relational test, made no errors (P1, P3, P15 and P19), while one participant made one error (P11), and one participant made two errors (P17). For the four participants (P5, P7, P9 and P13) from the All-More group that failed to meet test criterion, two participants (P5 and P7) made three errors, one participant (P13) made five errors, and one participant (P9) made no correct responses on the two-node relations. A summary of performance accuracy across each exposure to testing for participants in the All-Less group that passed and failed the experiment can be seen in Table 3.3. When exposed to the two-node relations, six participants from the All-Less group that met criterion during arbitrary relational testing, made no errors (P2, P4, P8, P10, P12, P16), two participants made one error (P14 and P20), and one participant made two errors (P6). For the one participant (P18) from the All-Less group that failed to meet test criterion, P18 made four errors on the two-node relations. In summary, high accuracy was observed for the All-Less group on the two-node relations, while accuracy was slightly lower for the All-More group on these same test relations.

Statistical Analyses

A McNemar test revealed no significant differences in accuracy scores between the baseline and mutually entailed relations (p = 1.000), the baseline and one-node relations (p = .125), and the baseline and two-node relations (p = .125), for the All-More group. Similarly, no significant differences were noted in accuracy between the mutually entailed and one- (p = .250) and two-node relations (p = .250), or the one- and two-node relations (p = 1.000), for the All-More group.

A McNemar test revealed no significant differences in accuracy scores between the baseline and mutually entailed relations (p = 1.000), the baseline and one-node relations (p = .500), and the baseline and two-node relations (p = .500), for the All-Less group. Similarly, no significant differences were noted in accuracy between the mutually entailed and one- (p = 1.000) and two-node relations (p = 1.000), or the one- and two-node relations (p = 1.000), for the All-Less group.

The results of Experiment 4 demonstrated that six out of ten participants in the All-More group (60% yield) and 9 out of 10 participants in the All-Less group (90% yield), met criterion during the arbitrary relational test. For these participants, high accuracy was observed on the baseline, mutually entailed, and one- and two-node relations. For example, a number of participants from both training groups, made no errors on any of the test relations (e.g., P2, P3 and P10). The remaining participants only made between one and five errors in total (e.g., P1, P4, P6, P8, P10, P11, P12, P14, P15, P16, P17, P19 and P20). Thus, for participants that passed the experimental task, high accuracy was observed on all test relations.

However, it must also be noted that four participants from the All-More group and one participant from the All-Less group failed to meet criterion during arbitrary relational testing, following four exposures to the experimental task. For these participants, accurate responding was maintained on the baseline relations, but was somewhat lower on the mutually entailed, and one- and two-node relations. For example, P18 made between 0 and 2 out of 8 correct responses on the mutually entailed relations, 6 out of 12 correct responses on the one-node relations, and 4 out of 8 correct responses on the two-node relations. Similarly, accuracy for P13 ranged between 3 and 5 out of 8 correct on the mutually entailed relations, between 4 and 5 out of 12 correct on the one-node relations, and between 3 and 7 out of 8 correct on the two-node relations. Thus, at times, response accuracy for participants that failed to meet test criterion was at, or below chance levels.

One potential reason for these findings may be that the contextual cues for MORE-THAN and LESS-THAN were not functioning during the arbitrary relational phases, for these participants. Thus, it may be possible that participants ignored the contextual cues and responded only on the basis of the two discriminative stimuli presented. For example, as accuracy was lower on the mutually entailed relations that involved a contextual cue that was *different to training,* it appeared that participants were responding similarly to training. That is, if participants were trained to select B not A (B+A-) in the presence of the MORE-THAN cue, then they also appeared to select B not A in the presence of the LESS-THAN cue, at testing. Similarly, response accuracy on the one- and two-node relations was, for a number of participants that failed to meet test criterion, at, or below chance levels (P7, P9, P13 and P18). Taken with findings from the mutually entailed relations, this appears to suggest that the contextual functions of MORE-THAN and LESS-THAN established during the non-arbitrary relational phases were not functioning for these participants during the arbitrary relational phases. Thus, Experiments 5A and 5B were designed in an attempt to address these issues, and an observing response and a variant of the simple-to-complex protocol were incorporated into the experimental design.

Experiment 5A

Experiment 5A incorporated an observing response during the arbitrary relational phases in order to address an issue arising from Experiment 4. For example, a number of participants in Experiment 4 failed to respond accurately to the mutually and combinatorially entailed relations, and it was questioned as to whether the contextual cues for MORE-THAN and LESS-THAN were functioning during the arbitrary relational phases. More specifically, a number of participants in Experiment 4 displayed accuracy on the mutually entailed relations that was below chance levels (e.g., P5, P13 and P18). For instance, P18 only made between 0 and 2 correct responses on the mutually entailed relations, while P13 only made between three and five correct responses. One potential method of assessing whether the contextual cues were functioning during the arbitrary relational phases is through the inclusion of an observing response. Typically in studies that have used observing response procedures, participants are presented with two or more schedules of reinforcement, where reinforcement is provided on one schedule, and non-reinforcement on the other (Escobar & Bruner, 2008, 2009; Lieving, Reilly, & Lattal, 2006). The observing response itself has no effect on primary reinforcement (Wycoff, 1952, 1969), but does produce the stimuli that are associated with the components of the reinforcement schedules. For example, Dube and McIlvane (1999) examined the potential facilitative effects of an observing response in decreasing stimulus over-selectivity, and increasing response accuracy, in three individuals with learning difficulties. During the initial delayed matching-to-sample task (DMTS: in which there is a delay between the presentation of the sample and comparison stimuli), accuracy scores revealed that participants could match one, but not both of the sample stimuli. An intervention aimed at improving accuracy was then introduced by means of a differential observing response procedure that required the observation and discrimination of both sample stimuli. The authors found that when the observing responses were prompted, participants displayed large improvements in their accuracy scores. However, findings also revealed that when the observing response was no longer prompted, accuracy returned to earlier low levels. Thus, the authors concluded that the differential observing response led to substantial reductions in stimulus over-selectivity.

In a study relevant to the current research, Reilly et al. (2005) included an observing response during only the arbitrary relational training phase. For example, the contextual cue for MORE-THAN appeared first onscreen and participants were required to press the "T" key on the computer keyboard to produce the two comparison stimuli. The authors included this observing response to ensure that participants were attending to the contextual cue presented, and responding to the comparison stimuli in accordance with the cue presented. Experiment 5A therefore adopted a similar approach in an attempt to determine whether the contextual cues for MORE-THAN and LESS-THAN were functionally relevant during the arbitrary relational training and test phases.

A further aim of Experiment 5A was to employ a variant of the simple-to-complex testing protocol in an attempt to examine the pre-requisites necessary for the emergence of derived comparative responding. For example, and as previously mentioned, findings from Experiment 4 revealed that a number of participants that failed to pass the arbitrary relational test, displayed low levels of accuracy on the mutually entailed relations. Thus, an important issue arising from these findings is whether accurate responding in accordance with the properties of bidirectional, mutually entailed relations, is necessary to facilitate responding to combinatorially entailed relations. To date, the simple-to-complex protocol has been employed to examine equivalence relations, in which probes for equivalence are presented once responding in accordance with symmetrical relations has been established (e.g., Adams et al., 1993; Fields et al., 2000; Smeets, Barnes-Holmes, & Striefel, 2006; Smeets, van Wijngaarden, Barnes-Holmes, & Cullinan, 2004). For example, using a matching-to-sample procedure (MTS), Adams et al. (1993) first exposed participants to training with AB, which was followed by probes for BA symmetry. Next, participants were trained on BC and tested for CB symmetry. Following test phases in which both BA and CB symmetrical relations were presented, participants were exposed to test trials that probed for AC transitivity. Equivalence was then assessed with the presentation of CA. A final block of test trials was also presented, in which baseline, symmetrical, transitive and equivalence probe trials were presented in a random order. When the results of the simple-to-complex testing protocol were compared against those in which a complex-to-simple protocol was employed (equivalence probes were presented first, followed by probes for symmetry), Adams et al. found that equivalence classes were formed more efficiently with the former testing protocol. In addition, the simple-to-complex testing protocol resulted in much less inter-subject variability, a finding that the authors attribute to the sequential presentation of probes for symmetry, transitivity and equivalence. Similarly, Smeets et al. (2004) also found that MTS procedures allow equivalence responding to emerge more readily when a simple-to-complex testing protocol is employed. In addition, Smeets et al. (2006) found that the simple-to-complex protocol is also successful in producing equivalence responding and equivalence in reversal, when employed alongside the precursor to the Relational Evaluation Procedure (pREP).

Thus, Experiment 5A adopted a variant of this testing protocol, in which probes for one- and two-node combinatorial entailment were only presented once participants had successfully passed tests for mutual entailment. So, for example, following training on four adjacent stimulus pairs, participants in Experiment 5A were exposed to a test block, in which mutually entailed relations were presented alongside baseline relations. If participants successfully responded in accordance with the properties of mutual entailment, they were then exposed to a second test phase that probed for combinatorial entailment. Again, the mutually entailed relations were presented alongside the baseline relations in this test block.

Method

Participants

Ten participants, one male and nine female, ranging in age from 19 to 22 years (M_{age} = 19.60, *SD* = .97), were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the task, and were randomly assigned to either the All-More or All-Less group at the start of the experiment.

Procedure

The procedure for Experiment 5A was similar to Experiment 4, except that an observation response, and an additional arbitrary relational test phase were included.

Phases 1A and 1B: Non-arbitrary Relational Training and Testing. Both phases were identical to Experiment 4.

Phases 2A and 2B: Arbitrary Relational Training and Testing. Similar to Experiment 4, the contextual cue again appeared first onscreen, but was now followed by an observation response where participants were required to press the spacebar in order for the two comparison stimuli to appear. The inclusion of this observation response was to assess whether the contextual functions of "More-than" and "Less-than" established during the non-arbitrary relational phases came to exert control over responding during the arbitrary relational training and testing phases.

Following arbitrary relational training, the first test phase (Test 1) was introduced and consisted of only baseline and mutually entailed relations. In order to complete Test 1 and progress to Test 2, participants were required to achieve a minimum of 12 out of 16 correct responses (i.e., 75% accuracy) across all baseline relations. Each of the four mutually entailed relations were presented four times each during this test phase, resulting in a total of 16 test trials (see Table 3.4). Participants were required to make a minimum of 3 out of 4 correct responses (i.e., 75% accuracy) on each individual mutually entailed test trial. If a participant failed to reach this mastery criterion, then they were re-exposed to the entire experimental task from the very beginning for a maximum of three further exposures.

Test 2 was identical to the arbitrary relational test phase in Experiment 4, which probed for the maintenance of the baseline relations alongside mutually entailed, and one- and two-node combinatorially entailed relations. Participants were required to achieve a minimum of 30 out of 36 correct responses (i.e., 83%) in order to complete this test phase (see Table 3.4). A failure to do so meant that participants were re-exposed to the entire experimental task again from the very beginning for a maximum of three more times.

A final feature of Experiment 5A was that all arbitrary stimulus sets (Sets 1-5) and contextual cues (Sets 6-10) were counterbalanced across participants.

Table 3.4
Training and Test Trials Received by the Two Groups in Experiment 5A.

Group	Relation Type	Test Trial Type					
		Test 1					
		Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
	ME	B>A	C>B	D>C	E>D		
All-More	Baseline	B>A	C>B	D>C	E>D		
	ME	A<B	B<C	C<D	D<E		
		Test 2					
		Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
All-Less	ME	B>A	C>B	D>C	E>D		
All-More	Baseline	B>A	C>B	D>C	E>D		
All-More	ME	A<B	B<C	C<D	D<E		
		Relations Common to All Groups					
	CE1	C>A	D>B	E>C	A<C	B<D	C<E
	CE2	D>A	E>B	A<D	B<E		

Note. "Baseline" refers to test trials involving directly trained relations, and the acronym ME, CE1, and CE2 refer to test trials for mutually entailed and one- and two-node combinatorially entailed relations respectively. The inequality symbols, < (LESS-THAN) and > (MORE-THAN), denote the contextual cue that was presented: This indicates which comparison should be "selected over" the other, with the reinforced comparison to the left, and the punished comparison to the right of the inequality symbol. It is important to note that the actual contextual cues used in the present study consisted of abstract visual images, and not the inequality symbols described here, which are used for the purposes of clarity.

Results and Discussion

Of the ten participants that began Experiment 5A, one participant (P3) failed to complete non-arbitrary relational training, while another two participants (P2 and P8; see Table 3.5) ended their participation in the study following a number of unsuccessful exposures to Test 1 (mutual entailment) and Test 2 (combinatorial entailment). Data for these three participants will be excluded from further analysis. Two participants (P5 and P7) failed to achieve the pre-determined Phase 2B (Arbitrary Relational Test 1) criterion following four exposures to this test phase. On the other hand, five participants (P1, P4, P6, P9 and P10) successfully completed the experimental task, with the number of exposures required to testing ranging between

1 and 3 (*M* = 2.00, *SD* = 2.70). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Table 3.5 displays the trials to criterion for the non-arbitrary and arbitrary relational training phases. Participants that passed required between 10 and 67 (*M* = 43.40, *SD* = 15.89) trials to complete non-arbitrary relational training (Phase 1) and between 12 and 516 (*M* = 199.20, *SD* = 193.05) trials to complete arbitrary relational training.

When exposed to Test 1 (Phase 2B: Arbitrary Relational Test 1), two participants (P1 and P9) made no errors on the baseline relations, while two participants (P4 and P10) made one error, and another (P4) made two errors. When exposed to the mutually entailed relations, four participants (P1, P4, P6 and P9) made no errors, while one participant (P10) made one error.

When exposed to Test 2 (Phase 2B: Arbitrary Relational Test 2), three participants (P4, P9 and P10) made no errors on the baseline relations, while one participant (P1) made one error, and another (P6) made two errors. When exposed to the mutually entailed relations, all five participants made no errors. During probes for one-node combinatorial entailment, three participants (P1, P6 and P9) made no errors, while one participant (P4) made one error, and another (P10) made two errors. When exposed to the two-node relations, two participants (P1 and P6) made no errors, while three participants made two errors (P4, P9 and P10).

Participants who failed to pass the arbitrary relational test

Two participants (P5 and P7) failed to meet the pre-determined criterion for the arbitrary relational Test 1 phase. P5 made 8 out of 16 correct responses across all four exposures to the baseline relations, and between 7 and 8 out of 16 correct responses on the mutually entailed relations. P7 made between 14 and 15 correct responses on the baseline relations, and between 0 and 2 correct responses on the mutually entailed relations, during his/her four exposures to this test phase.

Table 3.5
Individual data for participants in Experiment 5A.

Participant	Phase 1A: Non-arbitrary Relational Training	Phase 2A: Arbitrary Relational Training	Phase 2B: Arbitrary Relational Test 1 (mutual entailment) B	ME	Phase 2B: Arbitrary Relational Test 2 (combinatorial entailment) B	ME	CE1	CE2
1(All-More)	23	516	16/16	1/16				
	10	12	16/16	16/16	15/16	16/16	24/24	16/16
2(All-More)**	689	84	15/16	2/16				
	23	12	16/16	0/16				
	10	12	16/16	0/16				
	18							
3(All-More)**	26	364						
4(All-More)	33	180	15/16	10/16				
	10	36	15/16	16/16	16/16	16/16	23/24	15/16
5(All-More)*	13	192	8/16	8/16				
	18	24	8/16	7/16				
	10	12	8/16	8/16				
	13	12	8/16	7/16				
6(All-Less)	35	96	14/16	16/16	14/16	16/16	24/24	16/16
7(All-Less)*	43	420	14/16	2/16				
	10	48	15/16	1/16				
	32	24	15/16	0/16				
	14	12	15/16	0/16				
8(All-Less)**	20	96	11/16	14/16				
	10	24	16/16	16/16	16/16	15/16	17/24	15/16
	10	36	15/16	12/16				
	10	24	16/16	15/16				
9(All-Less)	61	24	13/16	15/16				
	10	36	16/16	16/16	16/16	16/16	24/24	15/16
10(All-Less)	15	72	15/16	16/16	14/16	13/16	21/24	12/16
	10	12	14/16	13/16				
	10	12	15/16	15/16	16/16	16/16	22/24	15/16

Note. Data is displayed for the number of trials required to achieve mastery criterion for the non-arbitrary and arbitrary relational training phases. Also shown are the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline, mutually entailed, and one- and two-node combinatorially entailed relations during Test 2. The acronym B, ME, CE1 and CE2 refer to test trials for baseline, mutually entailed and one- and two-node combinatorially entailed relations respectively. * refers to participants that failed to complete the experimental task. ** refers to participants that quit the experiment.

The results from Experiment 5A revealed that one participant (P6) passed both test phases on his or her first exposure to testing, and he/she displayed high levels of accuracy on all test relations. In addition, three participants (P1, P4 and P2) completed the experimental task on their second exposure to testing, again displaying high levels of accuracy on all test relations. P10 on the other hand required three exposures to testing to meet criterion, but displayed high levels of accuracy on all relations when he/she met test criterion.

As mentioned, a number of participants that passed required more than one exposure to training and testing to do so. Interestingly, findings revealed that accuracy on the mutually entailed relations was not at criterion performance for two participants (P1 and P4) on their first exposure to testing. For example, on his/her first exposure to the experimental task, P1 only made 1 out of 16 correct responses on the mutually entailed relations. However, on his/her second exposure to testing, P1 achieved criterion performance on the mutually entailed relations. Similarly, P4 did not meet test criterion on the mutually entailed relations on his/her first exposure to testing, but did on his/her second exposure. Both participants (P1 and P4) then displayed accurate performances on all test relations during Test 2. In addition, findings from the two participants that failed to pass the experiment revealed that both failed to respond in accordance with the properties of mutual entailment. For example, P5 made between 7 and 8 out of 16 correct responses on these relations, while P7 only made between 0 and 2 correct responses. These findings would therefore appear to suggest that responding in accordance with the properties of mutual entailment, may, facilitate responding to combinatorially entailed relations. However, it is difficult to definitively conclude this, as only participants that met test criterion were exposed to tests for combinatorial entailment. Therefore, it may be beneficial for future studies to incorporate control conditions in which one group of participants that fail tests for mutual entailment are exposed to tests for combinatorial entailment, while a second group are not. This in turn may help to more fully determine whether failure to respond in accordance with the properties of mutual entailment affects our ability to respond to combinatorially entailed relations.

Experiment 5A also sought to determine the potential utility of an observing response in generating greater stimulus control during the arbitrary relational phases. As mentioned, results from Experiment 4 suggested that participants may be ignoring the MORE-THAN and LESS-THAN contextual cues during the arbitrary relational phases. Thus, in Experiment 5A, when, for example, the MORE-THAN cue appeared onscreen, participants were prompted to press the spacebar to allow the comparison stimuli to appear. Results demonstrated that of the seven participants that completed Experiment 5A, five met criterion and two did not. However, it is difficult to assess the extent to which the observing response contributed to these findings, as all participants were exposed to it during the arbitrary relational phases. Therefore, in order to more fully determine the utility of the observing response in establishing effective stimulus control, future studies should seek to include control conditions in which one group of participants receive the observing response, while a second group do not.

Experiment 5B

Experiment 5B sought to again determine the utility of a variant of the simple-to-complex protocol and an observing response in generating arbitrary comparative responding. In addition, some minor changes were made to the training and testing protocol in order to determine some of the potential factors affecting stimulus control. For example, in Experiment 5B, participants were not re-exposed to non-arbitrary relational training and testing if they failed to reach criterion during either arbitrary relational test phase. Instead, participants were re-exposed to arbitrary relational training, followed again by arbitrary relational testing. In addition, during the second arbitrary relational test phase in which probes for combinatorial entailment were presented, the mutually entailed relations were now omitted. The reason for this was to try and keep the current variant of the simple-to-complex protocol as similar to the original, and thus, present separate test blocks involving mutually and combinatorially entailed relations.

Method

Participants

Eight participants, two male and six female, ranging in age from 20 to 48 years (M_{age} = 24.38, SD = 9.50), were recruited via the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the task, and were randomly assigned to either the All-More or All-Less group at the start of the experiment.

Procedure

The procedure for Experiment 5B was almost identical to that of Experiment 5A. However, participants were not re-exposed to *Phase 1: Non-arbitrary Relational Training and Testing* if they failed to reach mastery criterion on any exposure to either arbitrary relational test (i.e., Tests 1 and 2). Instead, they were re-exposed to *Phase 2: Arbitrary Relational Training* instead (see Table 3.6).

Table 3.6
Training and Test Trials received by the two groups in Experiment 5B.

Group	Relation Type	Test Trial Type					
		Test 1					
		Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
	ME	B>A	C>B	D>C	E>D		
All-More	Baseline	B>A	C>B	D>C	E>D		
	ME	A<B	B<C	C<D	D<E		
		Test 2					
		Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
All-More	Baseline	B>A	C>B	D>C	E>D		
		Relations Common to All Groups					
	CE1	C>A	D>B	E>C	A<C	B<D	C<E
	CE2	D>A	E>B	A<D	B<E		

Note. "Baseline" refers to test trials involving directly trained relations, and the acronym ME, CE1 and CE2 refer to test trials for mutually entailed and one- and two- node combinatorially entailed relations, respectively. The inequality symbols, < (LESS-THAN) and > (MORE-THAN), denote the contextual cue that was presented: This indicates which comparison should be "selected over" the other, with the reinforced comparison to the left, and the punished comparison to the right of the inequality symbol. It is important to note that the actual contextual cues used in the present study consisted of abstract visual images, and not the inequality symbols described here, which are used for the purposes of clarity.

In addition, the test criterion and relations presented during Tests 1 and 2 differed from those presented in Experiment 5A (see Table 3.6). The mastery criterion for Test 1 was the same as in Experiment 5A. During Test 2, the mutually entailed relations were omitted, with only the baseline and one- and two-node relations presented. The criterion for Test 2 also differed from that in Experiment 5A. Participants were required to make a minimum of 12 out of 16 correct responses (i.e., 75% accuracy) across all baseline relations. However, they were also required to make a minimum of 3 out of 4 correct responses (i.e., 75% correct) on each one- and two-node test trial. Finally, similar to Experiments 4 and 5A, participants were exposed to the experimental task for a maximum of four times.

Results and Discussion

Of the eight participants that began Experiment 5B, one participant (P7) failed to complete non-arbitrary relational training, while another participant (P5; see Table 3.7) ended their participation in the study following a number of unsuccessful exposures to Test 1 (mutual entailment) and Test 2 (combinatorial entailment). Data for these two participants will be excluded from further analysis. Three participants (P1, P2 and P3) failed to achieve the pre-determined Test 1 criterion following four exposures to this test phase, while two participants (P6 and P8) failed to meet the pre-determined Test 2 criterion. Thus, only one participant (P4) successfully completed the experimental task, and required only one exposure to testing to do so. Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

When exposed to testing, P4 made one error on both the baseline and mutually entailed relations during Test 1 (Phase 2B: Arbitrary Relational Test 1). When exposed to Test 2 (Phase 2B: Arbitrary Relational Test 2), P4 made no errors on the baseline relations, one error on the one-node relations and two errors on the two-node relations.

Participants who failed to pass the arbitrary relational test

A total of three participants (P1, P2 and P3) failed to meet the pre-determined criterion for the arbitrary relational Test 1 phase (mutual entailment). Across all four exposures to Test 1, P1 made between 14 and 16 out of 16 correct responses on the baseline relations, and between 0 and 3 out of 16 correct responses on the mutually entailed relations. P2 made between 14 and 16 correct responses on the baseline relations, and between 0 and 1, correct responses on the mutually entailed relations. In addition, P3 made between 15 and 16, correct responses on the baseline relations, and between 0 and 1, correct responses on the mutually entailed relations.

Table 3.7
Individual data for participants in Experiment 5B.

Participant	Phase1A: Non-arbitrary Relational Training	Phase 2A: Arbitrary Relational Training	Phase 2B: Arbitrary Relational Test 1 (mutual entailment) B	ME	Phase 2B: Arbitrary Relational Test 2 (combinatorial entailment) B	CE1	CE2
1(All-More)*	20	96	16/16	1/16			
		12	15/16	3/16			
		12	16/16	1/16			
		24	16/16	0/16			
2(All-More)*	285	48	14/16	0/16			
		12	16/16	1/16			
		24	16/16	0/16			
		24	16/16	1/16			
3(All-More)*	18	84	16/16	1/16			
		12	16/16	1/16			
		12	16/16	0/16			
		12	15/16	0/16			
4(All-More)	44	24	15/16	15/16	16/16	23/24	14/16
5(All-Less)**	190	72	9/16	10/16			
		24	16/16	16/16	14/16	10/24	11/16
6(All-Less)*	15	60	15/16	16/16	10/16	22/24	7/16
		36	15/16	16/16	15/16	24/24	8/16
		12	16/16	15/16	15/16	23/24	8/16
		24	16/16	16/16	14/16	22/24	8/16
7(All-Less)**	154						
	604						
8(All-Less)*	74	36	14/16	15/16	15/16	16/24	16/16
		24	16/16	16/16	14/16	16/24	15/16
		12	16/16	16/16	16/16	16/24	16/16
		12	16/16	16/16	15/16	16/24	16/16

Note. Data is displayed for the number of trials required to achieve mastery criterion for the non-arbitrary and arbitrary relational training phases. Also shown are the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline, and one- and two-node relations in Test 2. The acronym B, ME, CE1 and CE2 refer to test trials for baseline, mutually entailed and one- and two-node combinatorially entailed relations, respectively. * refers to the participants who failed to complete the experimental task. ** refers to participants that quit the experiment.

Two participants (P6 and P8) met criterion during Test 1, but failed to meet criterion during Test 2. Across all four exposures to Test 1, P6 made between 15 and 16 out of 16 correct responses on the baseline and mutually entailed relations. Across his/her four exposures to Test 2, P6 made between 10 and 15 out of 16 correct responses on the baseline relations, between 22 and 24 out of 24 correct responses on the one-node relations, and between 7 and 8 out of 16 correct responses on the two-node relations. Finally, across all four exposures to Test 1, P8 made between 14 and 16 correct responses on the baseline relations, and between 15 and 16 correct responses on the mutually entailed relations. Across his/her four exposures to Test 2, P8 made between 14 and 16 correct responses on the baseline relations, 16 out of 24 correct responses on the one-node relations, and between 15 and 16 correct responses on the two-node relations.

The results of Experiment 5B failed to replicate the findings from Experiment 5A, with only one participant (P4) successfully completing the experimental task. Two participants (P6 and P8) failed to meet Test 2 criterion (combinatorial entailment), while three participants (P1, P2 and P3) failed to meet criterion during Test 1 (mutual entailment). More specifically, and similar to findings from Experiments 4 and 5A, participants that failed to pass tests for mutual entailment appeared to respond similarly to training. For example, P1, P2 and P3 only made between 0 and 3 correct responses on the mutually entailed relations, and thus, it could be proposed that for these participants, the contextual cues for MORE-THAN and LESS-THAN were not functioning during the arbitrary relational phases. Thus, effective stimulus control during the arbitrary relational phases was absent for a number of participants in Experiment 5B.

One potential reason for these findings may be due, in part, to the fact that, in comparison to Experiments 4 and 5A, participants were not re-exposed to non-arbitrary relational training and testing if they initially failed to meet criterion during arbitrary relational testing. For instance, some researchers propose that responding in accordance with non-arbitrary relations is an essential precursor to responding to arbitrary relations (e.g., Berens & Hayes, 2007; Hayes et al., 2001a; Stewart & McElwee, 2009). Furthermore, a number of studies have found that converting weak arbitrary relations into non-arbitrary forms facilitates responding to arbitrary comparative relations (see Barnes-Holmes et al., 2004; Vitale et al., 2008). Therefore, a potential limitation to Experiment 5B was that participants were not re-exposed to the non-arbitrary relational phases if they failed to meet arbitrary relational test criterion. Future studies should therefore seek to take this into consideration.

General Discussion

The current chapter was concerned with examining the effectiveness of a novel behavioural account of TI-like behaviour in adult populations. Throughout Experiments 4-5B, performance accuracy on comparative "More-than" and "Less-than" relations was analysed for two training groups (All-More and All-Less). A second aim of the current chapter was to investigate the effectiveness of interventions by means of an observing response, and a variant of the simple-to-complex testing protocol, in facilitating the emergence of this behaviour (Experiments 5A and 5B).

Experiment 4 examined performance accuracy on a number of "More-than" and "Less-than" test problems for participants exposed to either an All-More or All-Less training group. That is, participants in the All-More group received training on only "More-than" relations during the arbitrary relational phase, whereas participants in the All-Less group received training on only "Less-than" relations. Both groups of participants were then exposed to the same arbitrary relational test. Findings revealed yields of 60% for the All-More group and 90% for the All-Less group. Thus, a larger number of participants from the All-Less group successfully completed the experimental task. Despite this, participants that passed from both training groups displayed highly accurate performances on the mutually and combinatorially entailed test relations. However, a number of participants failed to demonstrate the predicted arbitrary comparative performances, and thus, it was proposed that the contextual cues for MORE-THAN and LESS-THAN were not functioning for these participants during the arbitrary relational phases. For example, accuracy for these participants was low on the mutually entailed relations and the one- and two-node relations that were *different to training*. That is, findings from Experiment 4 revealed that if participants learned to select B not A (B+A-) in the presence of the MORE-THAN cue, then they also selected B not A in the presence of the LESS-THAN cue during testing. Experiments 5A and 5B were designed in an attempt to address this issue.

Experiments 5A and 5B were concerned with addressing the low accuracy observed on some of the test relations for participants in Experiment 4. As mentioned, it was proposed that failures in contextual or stimulus control may have accounted for these findings, and thus, Experiments 5A and 5B sought to incorporate an observing response and a variant of the simple-to-complex protocol in an attempt to address this issue. Results from Experiments 5A and 5B demonstrated that there were yields of 50% and 12.5 %, respectively. For participants that passed both experiments, accurate performances were observed on the mutually and combinatorially entailed relations (Experiment 5A: P1, P4, P6, P9 and P10; Experiment 5B: P4). In contrast, for participants that failed to pass the experimental tasks, weaknesses were primarily noted on their ability to respond in accordance with the properties of mutual entailment (Experiment 5A: P5 and P7; Experiment 5B: P1, P2 and P3). Thus, the observing response and the variant of the simple-to-complex testing protocol resulted in only limited improvements in stimulus control in Experiments 5A and 5B.

Same relation as trained and different relation to trained

An interesting finding across all experiments in Chapter 3 was that accuracy was low on both the mutually entailed, and one- and two-node combinatorially entailed relations that were *different to training*. As an example, one group of participants were trained on B>A, C>B, D>C and E>D in the presence of the MORE-THAN contextual cue. Novel test trials then involved a combination of both "More-than" and "Less-than" relations. So, the mutually entailed relations presented to the All-More group at test, were A<B, B<C, C<D and D<E, in the presence of the LESS-THAN contextual cue (*different to training*). One-node test relations consisted of C>A, D>B and E>C, in the presence of the MORE-THAN contextual cue (*same as training*), and A<C, B<D and C<E, in the presence of the LESS-THAN contextual cue (*different to training*). Similarly, the two-node relations consisted of D>A and E>B, in the presence of the MORE-THAN contextual cue (*same as training*), and A<D and B<E, in the presence of the LESS-THAN contextual cue (*different to training*).

Previous research findings (e.g., Munnelly et al., 2010) have similarly reported that accuracy is lower on the one- and two-node relations that are *different to training*, due to the non-linear nature of these relations. For example, linear relations, such as E>C (i.e., relations that are the *same as training*), are said to be easier to solve due to the fact that they may simply be solved by re-arranging E>D and D>C into E>D>C, to infer the correct answer. On the other hand, non-linear relations such as C<E (i.e., relations that are *different to training*), require a greater response effort in order to solve the task. That is, participants are required to again re-arrange E>D and D>C into E>D>C, and then convert this into C<D<E in order to arrive at the correct conclusion (e.g., Hunter, 1957).

This distinction may again be applied to the findings from the current chapter. However, low accuracy on the one- and two-node relations that were *different to training* was not observed for participants in the current chapter that met criterion during arbitrary relational testing. Furthermore, for participants that failed to meet test criterion, P18 (All-Less) in Experiment 4, only made six out of twelve correct responses on the one-node relations, and four out of eight correct responses on the two-node relations. Similarly, for a number of other participants in Experiment 4 (All-More: P5, P7, P9 and P13), accuracy on the one- and two-node relations was below chance levels. Thus, the low accuracy observed on both the mutually entailed, and one- and two-node relations that were *different to training*, does raise the issue as to whether the contextual functions of MORE-THAN and LESS-THAN established for two abstract images during the non-arbitrary relational phases, extended to the arbitrary relational phases. That is, were participants responding appropriately to the discriminative stimuli in the presence of the particular contextual cue, or was it possible that participants may in fact have ignored the contextual cue, and responded only in terms of the two discriminative stimuli presented. Findings from participants in Experiment 4 (All-More: P13; All-Less: P18), and indeed Experiments 5A (P5 and P7) and 5B (P1, P2 and P3), that failed to respond to in accordance with the properties of mutual entailment, would appear to suggest that the cues were not functionally relevant during the arbitrary relational phases, as participants displayed low levels of accuracy on the mutually entailed test relations. Thus, Experiments 5A and 5B sought to determine whether greater stimulus control over responding could be achieved by incorporating an observing response throughout the arbitrary relational phases.

Observing Response

Experiments 5A and 5B attempted to address findings from Experiment 4 that stimulus control over responding was lacking for a number of participants. For example, participants that did not meet test criterion in Experiment 4 displayed low levels of accuracy on the mutually, and combinatorially entailed relations, and thus, it was proposed that for a number of participants, the contextual cues for MORE-THAN and LESS-THAN were not functioning during the arbitrary relational training and testing phases. Thus, Experiments 5A and 5B sought to examine the potential facilitative effects of an observing response on arbitrary comparative responding. For example, during the arbitrary relational phases, the contextual cue for MORE-THAN appeared first onscreen, and participants were then prompted by a message onscreen to press the spacebar in order to reveal the two comparison stimuli. The same applied when the contextual cue for LESS-THAN was presented. The observing response was included on the basis of previous findings which found that response accuracy for three individuals with developmental delays improved following the inclusion of an observing response throughout the course of the experimental task (Dube & McIlvane, 1999). With respect to the current findings, in Experiment 5A, five out of ten participants (50% yield) successfully passed both test phases, while only one out of eight participants (12.5% yield) in Experiment 5B, met criterion on both test phases. Thus, it appears that the observing response had a facilitative effect for participants that passed Experiments 5A and 5B, but not for participants that failed to meet test criterion. Again, similar to the problems noted in Experiment 4, the contextual functions of MORE-THAN and LESS-THAN were not functioning during the arbitrary relational phases for participants that failed to meet test criterion, despite the implementation of the observing response. Furthermore, in order to more clearly determine the potential facilitative effects of the observing response, it may be necessary for future studies to manipulate its presence and absence. For example, Dube and McIlvane (1999) found that although the inclusion of an observing response led to the desired improvements in performances for three individuals with developmental delays, when it was removed, accuracy returned to earlier lower levels. Therefore, in order to more fully determine the facilitative effects of the observing response on the emergence of arbitrary comparative responding, future studies should seek to compare performance accuracy for a group of participants that are exposed to the observing response, with a group that are not.

Variant of the simple-to-complex testing protocol

A second aim of Experiments 5A and 5B was to examine the potential facilitative effects of a variant of the simple-to-complex protocol on the emergence of arbitrary comparative responding. As mentioned earlier, response accuracy on the mutually entailed and one- and two-node relations that were *different to training* was low for a number of participants, and it was questioned as to whether responding in accordance with the properties of mutual entailment is necessary to facilitate responding to combinatorially entailed relations. Thus, in addition to the observing response, Experiments 5A and 5B sought to explore this by incorporating a variant of the simple-to-complex protocol. With this protocol, test relations were presented in a sequential order, where the mutually entailed relations were presented before tests for combinatorial entailment. Findings from the literature on equivalence support the utility of the simple-to-complex testing protocol in facilitating responding (e.g., Adams et al., 1993; Fields et al., 2000; Smeets et al., 2004), however, previously, studies examining derived comparative responding with adult participants, had not employed this testing protocol. In turn, employing the simple-to-complex protocol may be useful for future studies seeking to examine the pre-requisites necessary for the emergence of arbitrary comparative responding relations in young children.

With respect to Experiments 5A and 5B, it is difficult to assess the potential contribution of the variant of the simple-to-complex protocol on the current findings as it was employed alongside the observing response intervention in both experiments. However and as previously mentioned, only limited improvements in performance accuracy were noted for a number of participants in Experiments 5A and 5B over those observed Experiment 4. Thus, in order to determine the potential utility of the variant of the protocol on performance accuracy for participants that passed Experiments 5A and 5B, future studies should seek to examine both interventions separately. Furthermore, a potential limitation to Experiments 5A and 5B is that a comparison of different arbitrary relational test protocols was not undertaken. For example, previous studies examining the effectiveness of the simple-to-complex testing protocol in generating equivalence relations have compared test performances for these participants, against a group exposed to a complex-to-simple protocol (e.g., equivalence probes are presented before symmetry probes; Adams et al., 1993). In one such study, Adams et al. (1993) found that equivalence classes were formed more efficiently with the simple-to-complex protocol, and much less inter-subject variability was observed, a finding which the authors attribute to the sequential presentation of probes for symmetry, transitivity and equivalence.

Therefore, future studies seeking to examine the effectiveness of this protocol in establishing arbitrary comparative responding in adult humans, should seek to undertake such a comparison.

Alternative interventions

As mentioned, findings from Experiments 5A and 5B revealed that, for a number of participants, effective stimulus control was still lacking despite the implementation of a number of interventions (e.g., observing response and variant of the simple-to-complex protocol). Moreover, the finding that both interventions did not lead to the predicted improvements in bidirectional responding highlights the fact that alternative interventions need to be developed and implemented, if necessary. For example, participants in Experiment 5B were not re-exposed to non-arbitrary relational training and testing if they failed to reach criterion on any exposure to arbitrary relational testing, whereas participants in Experiment 5A were. Indeed, a number of studies have highlighted the importance of a history of responding in accordance with the non-arbitrary properties of stimuli, as a precursor to the ability to respond in accordance with arbitrary comparative relations (e.g., Berens & Hayes, 2007; Gorham et al., 2009; Vitale et al., 2008). For example, Vitale et al. (2008) employed an intervention consisting of multiple-exemplar training (MET) with non-arbitrary stimulus sets when adult participants displayed weak performances on a number of "More-than" and "Less-than" problem-solving tasks. The intervention consisted of converting the deficient arbitrary relations into a non-arbitrary form, and the authors reported that the intervention had a facilitative effect on problem-solving abilities at test. With respect to the current study, it may have been beneficial to expose the participants that initially displayed weak relational performances, to an intervention by means of non-arbitrary relational training. This in turn may have led to improvements in performances, and thus, future studies should seek to incorporate such interventions, when necessary.

The development of alternative training and testing protocols

One issue that has arisen from the current chapter is that, it may be possible that the conditional discrimination paradigm itself, contributed to the low accuracy observed for participants, on a number of test relations. For example, participants were exposed to a selection-based task, in which they were required to press a number of designated keys on the computer keyboard in order to make their response. However, it has been proposed that selection-based response systems such as those employed in conditional discriminations, involve multiple response forms (e.g., scanning, selecting and/or handing over; Sundberg & Michael, 2001), and conditional discrimination paradigms that are often more complex than they first appear (e.g., Lowenkron, 1991; Michael, 1985; Sundberg & Michael, 2001; Sundberg & Sundberg, 1990). This in turn is an important issue if such procedures are to be adapted to examine the emergence of derived comparative responding in young children, and children with autism who may lack the ability to make responses using a number of different modalities (e.g., Sundberg & Michael, 2001). Furthermore, when participants make responses to conditional discriminations, they are not provided with an opportunity to evaluate their selection before the next trial commences. However, the Relational Completion Procedure (RCP; Dymond & Whelan, 2010) originally developed to examine "Same" and "Opposite" relations, offers a more evaluative-based alternative for examining the emergence of derived stimulus relations. Chapter 4 will therefore incorporate a variant of the RCP, to determine its utility in generating more accurate responding to arbitrarily applicable comparative relations.

In conclusion, findings from Experiment 4-5B demonstrate that responding to a combination of arbitrary comparative "More-than" and "Less-than" relations, has the potential to provide a novel behaviour-analytic account of TI in adult participants. In addition, the current chapter sought to determine the utility of a number of interventions (e.g., variant of the simple-to-complex protocol and an observing response), in generating accurate performances at test. Findings revealed that despite the implementation of these interventions, only limited improvements in response accuracy were noted across Experiments 5A and 5B, and thus, further research is needed to determine the factors influencing the emergence of arbitrary comparative responding. More specifically, across all experiments in the current chapter, weaknesses were noted for a number of participants in their ability to respond in accordance with the properties of mutual entailment. Chapter 4 will therefore attempt to further explore the conditions necessary to facilitate the emergence of this pattern of responding. In addition, Chapter 4 will seek to determine the utility of a novel training and testing protocol, the RCP, in generating these performances.

Chapter 4

Establishing Arbitrarily Applicable Comparative Responding with the Relational Completion Procedure

The experiments reported in the current thesis have to date employed the conditional discrimination to examine the emergence of arbitrarily applicable comparative responding. The protocol involves the presentation of two conditional stimuli (samples) and two or more discriminative stimuli (comparisons). For instance, in a study on establishing responding in accordance with the relational frames of "More-than" and "Less-than", Reilly et al. (2005) compared response latencies to arbitrary test relations for three training groups; All-More, All-Less and Less-More. Participants were first exposed to non-arbitrary relational training and testing, followed by arbitrary relational training and testing, using procedures similar to those described in Chapter 3. Results demonstrated that response latencies were significantly faster on all test relations for participants in the All-More group, in comparison to the other two groups (see also Munnelly et al., 2010).

As mentioned, in the Reilly et al. (2005) study, the authors employed the conditional discrimination paradigm to examine the emergence of derived comparative responding. In addition, in order for participants to respond to conditional discriminations in this study, Reilly et al. (2005) incorporated a selection-based response system. For example, selection-based response systems are defined as a type of verbal behaviour in which an individual points to, touches, or selects a stimulus, from an array of stimuli involved in a conditional discrimination (e.g., Michael, 1985; Shafer, 1993; Vignes, 2007). Indeed, these procedures have been employed in the form of communication boards in studies involving young children with verbal impairments (e.g., Adkins & Axelrod, 2001; Potter & Brown, 1997). For example, the communication board consists of several pictures on one board, where the child must point to the item or activity that they request in order to gain access to it (Adkins & Aexlrod, 2001). With respect to the Reilly et al. (2005) study, the selection-based response system consisted of participants pressing the "z" or "m" keys, to select the comparison stimulus on the left or right-hand side of the computer screen, respectively.

Despite the fact that the conditional discrimination is the primary training and testing protocol employed in studies examining multiple stimulus relations, other protocols employed to examine equivalence relations have allowed for variations of the conditional discrimination to be developed. For example, the constructed-response matching-to-sample (CRMTS) protocol has been employed to study the emergence of spelling repertoires in both typically developing individuals, and individuals with developmental delays (e.g., deSouza, Goyos, Silvares, & Saunders, 2007; Dube, McDonald, McIlvane, & Mackay, 1991). With this protocol, participants are required to "construct" the correct comparison stimulus from its individual components (e.g., deSouza et al., 2007; Dube et al., 1991). For instance, a participant may be presented with the written word "cat" as a sample stimulus, and selecting the letters "c", "a", and "t", from a choice pool, would indicate the correct comparison stimulus. Dube et al. (1991) used a CRMTS procedure to establish spelling repertoires in two individuals with developmental delays and found it was effective in establishing generalized identity matching for both individuals.

Although there are many advantages associated with the conditional discrimination and MTS, some limitations have been noted. For instance, the top-down method of presenting the contextual cue above both comparison stimuli does not reflect the order in which individuals encounter relational stimuli in their everyday environment. Typically, when individuals engage in tasks, such as reading non-Arabic languages (e.g., English), the stimuli are presented (read) in sequence from left-to-right. Indeed, Mackay and Fields (2009) propose that non-arbitrary properties, such as the position of events in sequences, are critical for performances on learning tasks. Thus, it may be beneficial for researchers examining the emergence of multiple stimulus relations to incorporate training and testing procedures that are more reflective of real-life experiences. Indeed, a number of alternative training and testing paradigms have been developed to examine multiple stimulus relations, other than equivalence.

One such alternative is the Relational Evaluation Procedure (REP), which has been employed to examine "Same" and "Different" (Stewart, Barnes-Holmes, & Roche, 2004), and "Before" and "After" (O'Hora et al., 2004) relations. Another alternative lies in the Relational Completion Procedure (RCP; Dymond & Whelan, 2010; see also Dymond, Ng, & Whelan, 2013), which has incorporated a novel approach to examine "Same" and "Opposite" relations. With respect to the REP, participants are required to evaluate the relation between different stimuli that are presented on a given trial. Thus, the key feature of this paradigm is that participants confirm or deny the applicability of particular stimulus relations to other sets of stimuli. For example, participants may be presented with a contextual cue for "Different" and two arbitrary images that are specified as being different from one another. The arbitrary images have previously been established as meaning TRUE or FALSE, and participants must then choose between these two images. In this instance, selecting TRUE in the presence of the DIFFERENT cue, and FALSE in the presence of the SAME cue, would be reinforced (Stewart et al., 2004).

The RCP on the other hand, developed to examine "Same" and "Opposite" relations, involves the sequential presentation of stimuli from left-to-right on the computer screen. For example, participants are presented with a sample stimulus, contextual cue and a blank comparison on the top half of the screen, followed by three comparison stimuli on the bottom half of the screen. Participants are required to "drag-and-drop" one of the comparison stimuli from the bottom of the screen to the blank comparison on the top of the screen. Similar to the REP, participants are required to evaluate their responses by confirming their selection via one of two buttons at the bottom of the screen, "Finish Trial" or "Start Again". Dymond and Whelan (2010) found a facilitative effect for the confirmatory response requirement and a greater number of participants successfully completing the experimental task following training and testing with the RCP, in comparison to those that received a MTS training and testing protocol. On the basis of these findings, the authors suggest that the RCP may hold utility as a procedure for training and testing other types of multiple stimulus relations. In addition, the RCP represents an attempt to move beyond standard MTS procedures, and provides participants with the opportunity to engage in a more evaluative form of responding (Hayes & Barnes, 1997), the facilitative effects of which remain to be determined.

Current Experiments

The current experiments adopted a variation of the RCP in an attempt to extend this research to "More-than" and "Less-than" relations. Experiments 6A and 6B were concerned with investigating the utility of the RCP in establishing responding to a 5-member relational network. A second aim of these experiments was to examine the effectiveness of a constructed-response training and testing format in establishing responding to non-arbitrary and arbitrary comparative relations. Thus, in Experiments

6A and 6B, participants were first exposed to non-arbitrary relational training and testing, followed by constructed-response non-arbitrary and arbitrary relational training and testing. During the arbitrary relational training phase, participants were exposed to one of three training designs, in which the training pairs differed between the groups. Participants in the All-More group were trained on the following four stimulus pairings: B>A, C>B, D>C and E>D in the presence of the MORE-THAN contextual cue; the All-Less group on A<B, B<C, C<D and D<E in the presence of the LESS-THAN contextual cue; and the Less-More group on A<B and B<C in the presence of the LESS-THAN contextual cue, and D>C and E>D in the presence of the MORE-THAN contextual cue (it is important to note that ">" and "<" are used here to denote the contextual cues of MORE-THAN and LESS-THAN, respectively. Participants were not exposed to these inequality symbols, but instead, two abstract visual images as contextual cues). At testing, all participants were exposed to probes for both mutual and combinatorial entailment.

Thirdly, the current chapter was concerned with addressing the lack of stimulus control observed across a number of experiments in Chapter 3. For example, and as proposed in the General Discussion of Chapter 3, at times, participant responding failed to come under effective stimulus control. This observation was most notable for participants that failed to meet criterion during arbitrary relational testing. For instance, it was proposed that if participants learned to select B not A (B+A-) in the presence of the MORE-THAN cue during arbitrary relational training, then they also selected B not A in the presence of the LESS-THAN cue at testing. Thus, for a number of participants in Chapter 3, the contextual cues for MORE-THAN and LESS-THAN were not functionally relevant during the arbitrary relational phases. Furthermore, the implementation of a number of interventions in Chapter 3 (repeated exposure to training and test phases, an observing response and a variant of the simple-to-complex protocol), only generated limited improvements in response accuracy. In turn, this led to the question being posed as to whether the conditional discrimination may have contributed to the failures in stimulus control. For instance, and as mentioned, in a conditional discrimination, stimuli are presented in a top-down manner, which is contrary to the method in which individuals encounter stimuli in everyday tasks, such as reading and writing (Dymond & Whelan, 2010). Thus, all experiments in Chapter 4 employed a variant of the RCP, in order to determine whether this protocol may overcome some of the problems with stimulus control encountered in Chapter 3.

Lastly, all experiments in Chapter 4 sought to examine the effects of the linearity of training pairs on relational performances at test. Typically, training pairs are presented in either a linear (i.e., sequential) or non-linear (i.e., random) order. For example, linear training relations involve presenting the trials in a sequential order (e.g., A<B, B<C, C<D and D<E), while non-linear relations involve presenting the trials in a non-sequential, random order (e.g., C<D, A<B, D<E and B<C). Some researchers suggest that when training pairs are presented in a non-linear order as opposed to a linear order, test problems are more difficult to solve (e.g., Hunter, 1957; Russell, McCormack, & Lillis, 1996). Hunter (1957) argues that because the stimulus pairs are presented in a non-linear order during training, participants must convert these pairs to a linear order before they can solve the test problems.

Two previous studies conducted under the rubric of RFT have investigated this issue (Gorham et al., 2009; Vitale et al., 2008). Gorham et al. (2009) exposed five typically developing children and three children with autism to "More-than" and "Less-than" arbitrary relations involving 2, 3, 4, or 5 identically sized coins. All children initially failed baseline tests involving the targeted (A-B, B-C, A-B-C, and A-B-C-D) and transitive relations (B-D), and thus, were exposed to a training phase in which these relations were presented in a linear order. Results from both experiments demonstrated that seven participants met training criterion and passed a subsequent test phase for the targeted and transitive relations when they were presented with novel stimulus sets. In addition, Vitale et al. (2008) examined the effects of linearity on arbitrary comparative test performances with a 3-term series task. According to the definition of linearity adopted in the current thesis, Vitale et al. exposed participants to tasks involving specified-same (linear: A>B and B>C or A<B and B<C) and specified-same transitive (non-linear: A>B and C>A, or A<B and C<A) arbitrary relational training relations. In addition, participants were exposed to a number of other training tasks involving unspecified relations (e.g., A>B and C>B). Irrespective of whether the trials were presented in a linear or non-linear order, all participants were exposed to the same arbitrary relational test. Across five experiments, the authors found that accuracy was comparably high on 3-term linear and non-linear arbitrary relational tests.

Experiments 6A and 6B therefore, aimed to examine the effects of linearity on performance accuracy at testing to a 5-member relational network. Participants in Experiment 6A were exposed to the training pairs in a linear order, while participants in Experiment 6B were exposed to the training pairs in a non-linear order.

Experiments 7A-8B were also concerned with the effectiveness of the RCP, and linearity, in generating arbitrarily applicable comparative responding.

Experiment 6A

Method

Participants

Twelve participants, three male and nine female, ranging in age from 18 to 30 years (M_{age} = 21.42, SD = 4.40), were recruited through personal contacts and the psychology subject pool at Swansea University. Participants were allocated partial course credit on completion of the task, and were randomly assigned to one of three training groups at the start of the experiment (All-More, All-Less or Less-More).

Apparatus and Setting

The experiment was conducted in an experimental room (2 X 3 metres) in the Psychology Department at Swansea University. All training and test trials were presented on a 16-inch display screen by a programme written in Visual Basic.NET ™.

Materials and Stimuli

Two arbitrary images were randomly selected from the Windings font, and were employed as contextual cues for MORE-THAN and LESS-THAN (see Figure 4.1) during the non-arbitrary and arbitrary relational training and testing phases. Twenty-eight stimulus sets were employed during non-arbitrary relational training and testing, and consisted of images of varying quantities of particular objects (see Appendix 5 for a list of the non-arbitrary stimulus sets employed in Experiments 6A-8B).

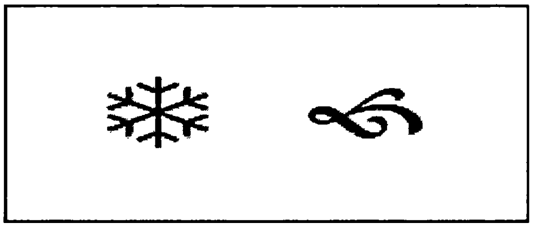


Figure 4.1. The two arbitrary stimuli employed as MORE-THAN and LESS-THAN contextual cues.

For the arbitrary relational training and testing phases, five images selected from the Kanji script (see Figure 4.2) were used to generate a 5-member linear relational network.

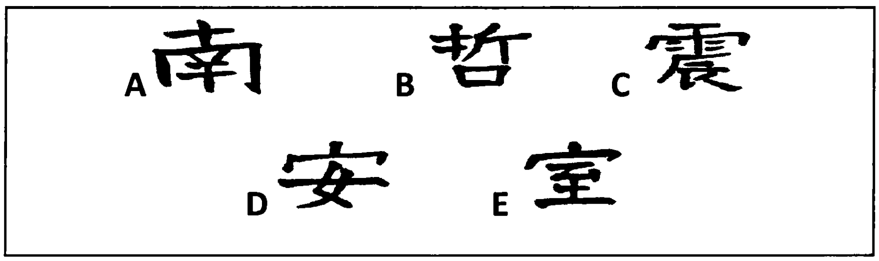


Figure 4.2. The five Kanji images employed during arbitrary relational training and testing, and labelled A, B, C, D, and E (Note: participants were never exposed to these labels).

Procedure

The general procedural sequence was based on those employed by Dymond and Whelan (2010), and was as follows: *Phase 1A: Non-arbitrary Relational Training* and *Phase 1B: Non-arbitrary Relational Testing; Phase 2A: Constructed-Response Non-arbitrary Relational Training* and *Phase 2B: Constructed-Response Non-arbitrary Relational Testing; Phase 3: Constructed-Response Arbitrary Relational Training; Phase 4: Arbitrary Relational Test 1;* and *Phase 5: Arbitrary Relational Test 2.*

During all training and testing phases, the computer screen was separated in two, the top two thirds of the screen was blue, whilst the bottom third was white. During Phase 1, the sample stimulus appeared first in the upper left-hand side of the screen. Following a delay of 1 s, the contextual cue (e.g., MORE-THAN/LESS-THAN) appeared in the upper centre of the screen, and a blank yellow square was presented following a 1 s delay in the upper right-hand side of the screen. Following a further delay of 1 s, two comparison stimuli appeared simultaneously on the lower third of the screen. The screen position (i.e., left or right) of these comparisons was counterbalanced across trials.

In order to make a response, participants were required to “drag” one of the two comparison stimuli and “drop” it in the blank yellow square. This was done by placing the mouse cursor over the comparison stimulus, they had selected as their response. Immediately upon making this selection, a red border appeared around the comparison stimulus to highlight their selection. Participants then clicked on, and held down the left mouse button, whilst dragging their selection to the blank yellow square. Releasing the left mouse button allowed the selected comparison to “drop” into the blank yellow square. At the same time, the screen position in which the comparison stimulus has originally appeared was replaced by a blank yellow square.

Once the selected comparison was placed in the blank yellow square, two confirmatory response buttons appeared simultaneously at the bottom of the screen. One button was labelled “Finish Trial”, and by hovering the mouse cursor over this button, a small text box with the caption “Click here to Finish Trial” appeared onscreen. The second button was labelled “Start Again”, and hovering over this button produced the caption “Click here to Start Again”. If the participant pressed the “Start Again” button, this cancelled the selection and resulted in all stimuli returning to their original positions before the selection was made. That is, the comparison stimulus that was selected returned to either the lower left or right portion of the screen, and the blank yellow square returned to the upper right of the screen.

All stimuli remained onscreen until the participant pressed the “Finish Trial” button. Training trials were then followed by feedback presented on a blue background for the duration of 3 s. When a participant made a correct response, feedback consisted of, from left-to-right, the sample, contextual cue, and the comparison stimulus the participant had selected on the previous trial. A yellow border surrounded all three images, and the word “Correct!” was presented in black underneath. A brief audible beep was presented following the word “Correct!” The only difference between feedback for a correct selection, and feedback for an incorrect selection, was that the word "Correct!" was now replaced by the word "Wrong", and no audible tone followed feedback. During all testing trials, no feedback was presented, and instead, the screen cleared and remained blue for the duration of 3 s. An inter-trial interval (ITI) of 1 s followed each trial, where the computer screen cleared and remained blue for the duration of the ITI.

The presentation of stimuli differed slightly during Phases 2-5. For example, the sample stimulus in the upper left-hand side of the screen was now replaced with a blank yellow square. Therefore, participants were presented with a blank yellow square, followed by a contextual cue, and another blank yellow square in the upper portion of the screen. Similar to Phase 1, two comparison stimuli were again presented on the lower portion of the screen.

During these phases, participants were required to "construct" their responses, from left-to-right in the upper portion of the computer screen. That is, participants were instructed to place one of the comparison stimuli in the upper-left blank yellow square, and the other comparison in the upper-right blank yellow square. Both the initial response, and confirmatory response requirements were identical to Phase 1. Again, all training trials were followed by feedback, whereas feedback was omitted during all test phases.

Feedback Thermometer. A task feedback thermometer was displayed in the centre, right-hand side of the screen during all training and testing phases (Fienup, Covey, & Critchfield, 2010). During training, the thermometer displayed the mastery criterion needed to complete training (e.g., "You need this many correct to move on: 10"), the current number of correct responses (e.g., 6 out of 10), and was incremented following every correct response. During testing, the thermometer displayed the total number of trials in the particular test phase, the current trial number, and incremented following every response.

Phase 1A: Non-arbitrary Relational Training. This phase began with the following instructions onscreen:

> Thank you for agreeing to participate in this study. You will be presented with a series of images on the top half of the screen from left

to right. Then you will be presented with 2 images on the bottom of the screen. Your task is to observe the images that appear from left to right and place one of these images from the bottom to the blank, yellow square. To select the image on the bottom, click on it once, and to place it in the blank square, click on this once. To confirm your choice, then click "Finish Trial". If you wish to make another choice, then click "Start Again". Sometimes you will receive feedback on your choices, but at other times you will not. Your aim is to get as many tasks correct as possible. It is always possible to get a task correct, even if you are not given feedback. If you have any questions please ask the experimenter. Please press the OK button below to begin the experiment!

Clicking on the OK button removed the instructions and signalled the start of Phase 1A. This training phase aimed to establish contextual control over responding for two arbitrary visual images (MORE-THAN and LESS-THAN) to comparison stimuli of varying quantities. So, for example, on a given trial, participants were presented with a sample (e.g., two basketballs), a contextual cue (e.g., MORE-THAN) and a blank yellow square in the upper portion of the screen. Two comparison stimuli (e.g., one and four basketballs) were also presented on the lower portion of the screen. In this instance, placing the comparison stimulus containing one basketball in the blank yellow square counted as a correct response. On the other hand, if two basketballs were again presented as the sample, alongside the contextual cue for LESS-THAN, and one and four basketballs as comparison stimuli, placing the comparison stimulus containing four basketballs in the blank yellow square was reinforced (see Figure 4.3). All training trials were followed by feedback presented for the duration of 3 s, and this was followed by an ITI of 1 s.

Four stimulus sets were employed during non-arbitrary relational training. Mastery criterion for this training phase was set at 10 consecutive correct responses. If participants met this criterion, they immediately proceeded to the non-arbitrary relational test phase. However, if participants failed to meet this criterion following exposure to 240 training trials, they were then exposed to a second non-arbitrary relational training phase, which involved four novel stimulus sets.

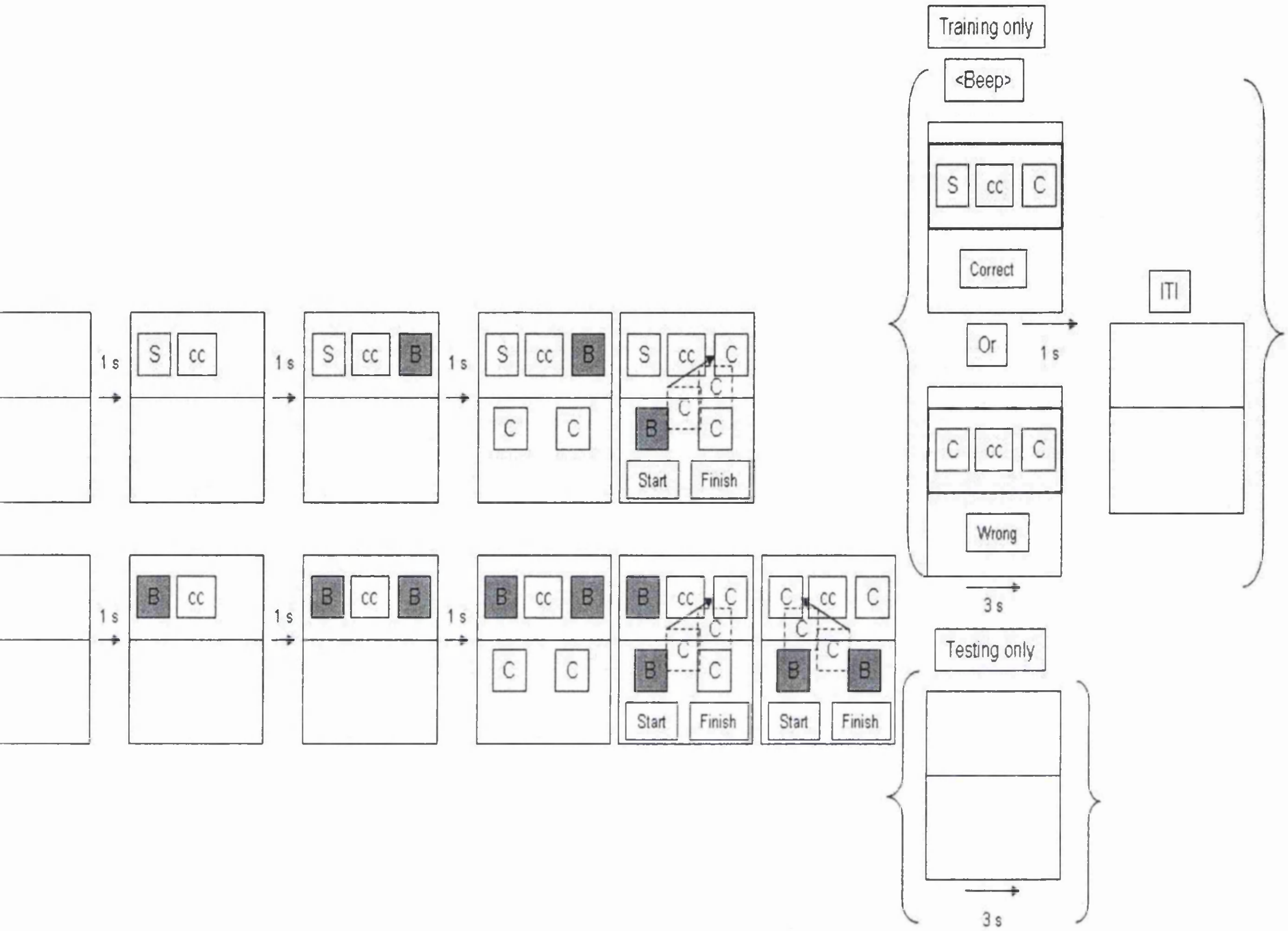


Figure 4.3. Schematic diagram of the sequence of presentation of stimuli during the non-arbitrary relational training and testing phases (upper panel) and the constructed-response non-arbitrary and arbitrary relational training and test phases (lower panel). *Note.* S = sample, cc = contextual cue, B = blank square, C = comparison, and a dashed line represents "dragging" a comparison stimulus. The text, "Finish" and "Start" represent the confirmatory response buttons. Arrows pointing from B to C illustrate that, once selected, the comparison stimulus moved to the upper portion of the screen, while the screen position in which it was originally, was now replaced by a blank square.

Phase 1B: Non-Arbitrary Relational Test. This phase was again similar to Phase 1A. However, in this test phase, participants were now presented with four novel stimulus sets, and all feedback was omitted. Participants were presented with a total of eight test trials and were required to respond correctly across all test trials in order to progress to the next phase of the experiment. However, if participants failed to meet this criterion, they were re-exposed to non-arbitrary relational training (i.e., Phase 1A) involving the same four stimulus sets. This was again followed by a non-arbitrary relational test.

Phase 2A: Constructed-Response Non-arbitrary Relational Training. This phase of the experiment began with the following instructions onscreen:

> The first phase of the experiment is now finished. You will now be presented with two blank yellow squares in the top left- and right-hand sides of the screen, and one image in the centre top of the screen. Then you will be presented with two images on the bottom of the screen. Your task is to drag and drop one image at a time from the bottom of the screen into the blank yellow squares. You must drag-and-drop an image into the left-hand blank yellow square and then drag-and-drop the next image into the right-hand yellow square. To select the image on the bottom, click on it once, and to place it in the blank square, click on this once. To confirm your choice, then click "Finish Trial". If you wish to make another choice, then click "Start Again". Sometimes you will receive feedback on your choices, but at other times you will not. Your aim is to get as many tasks correct as possible. It is always possible to get a task correct, even if you are not given feedback. If you have any questions please ask the experimenter. Please press the OK button below to continue!

The purpose of this phase was to train participants to "construct" the relation between two comparison stimuli, in the presence of a particular contextual cue. On each trial, participants were presented with a blank yellow square, a contextual cue, and another blank yellow square in the upper portion of the screen. Participants were also presented with two comparison stimuli on the lower portion of the screen. So, for example, participants were presented with the contextual cue for MORE-THAN, and six and four guitars as the comparison stimuli. A correct response consisted of "dragging" and "dropping" the six guitars to the upper-left blank yellow square and the four guitars to the upper-right blank yellow square, *in that sequence*. Similarly, if six guitars and four guitars were presented as comparison stimuli in the presence of the LESS-THAN contextual cue, placing the four guitars in the upper-left blank yellow square, and the six guitars in the upper-right blank yellow was reinforced. Again, feedback was presented following all training trials.

Participants were presented with four stimulus sets during training, and mastery criterion was set at 10 consecutive correct responses. If participants successfully met training criterion, they were immediately exposed to the constructed-response non-arbitrary relational test phase. However, similar to Phase 1A, if participants were unsuccessful in meeting training criterion after exposure to 240 training trials, they were re-exposed to constructed-response non-arbitrary relational training, followed again by constructed-response non-arbitrary relational testing.

Phase 2B: Constructed-Response Non-Arbitrary Relational Testing. This phase was identical to Phase 2A, with the exception that participants were presented with four novel stimulus sets, and feedback was no longer provided. Participants were exposed to eight test trials, and were required to respond correctly across all test trials to progress to the next phase of the experiment. If participants failed to meet this criterion, they were re-exposed to constructed-response non-arbitrary relational training involving the same four stimulus sets. This was again followed by a constructed-response non-arbitrary relational test phase.

Phase 3: Constructed-Response Arbitrary Relational Training. This phase commenced immediately upon completion of Phase 2B, with the following instructions onscreen:

> The second phase of the experiment is now finished. You will again be presented with two blank yellow squares in the top left- and right-hand sides of the screen, and one image in the centre top of the screen. Then you will be presented with two images on the bottom of the screen. Your task is to drag and drop one image at a time from the bottom of the screen into the blank yellow squares. You must drag-and-drop an image into the left-hand blank yellow square and then drag-and-drop the next image into the right-hand yellow square. To select the image on the bottom, click on it once, and to place it in the blank square, click on this once. To confirm your choice, then click "Finish Trial". If you wish to make another choice, then click "Start Again". Sometimes you

> will receive feedback on your choices, but at other times you will not. Your aim is to get as many tasks correct as possible. It is always possible to get a task correct, even if you are not given feedback. If you have any questions please ask the experimenter. Please press the OK button below to continue!

Similar to Phase 2, participants were presented with a blank yellow square, a contextual cue, and another blank yellow square in the upper portion of the screen. Again, two comparison stimuli were presented simultaneously on the lower portion of the screen. However, during this phase, the comparison stimuli consisted of arbitrary images randomly selected from the Kanji script, which are labelled A, B, C, D, and E (see Figure 4.2).

Participants were presented with training trials in a linear order, and training pairs differed between the All-More, All-Less, and Less-More training groups. The All-More group were trained B>A, C>B, D>C and E>D, in the presence of the MORE-THAN contextual cue; the All-Less group A<B, B<C, C<D and D<E, in the presence of the LESS-THAN contextual cue; and the Less-More group A<B and B<C, in the presence of the LESS-THAN contextual cue, and D>C and E>D, in the presence of the MORE-THAN contextual cue. All training pairs were presented in this order for all three groups (see Figure 4.4. and Table 4.1).

For all groups, the four training pairs were presented for a total of three times each, resulting in a block of 12 training trials. Mastery criterion for the arbitrary relational training phase was set at 12 out of 12 correct responses (i.e., 100% accuracy) on any given block. Training blocks were repeated until participants achieved this criterion.

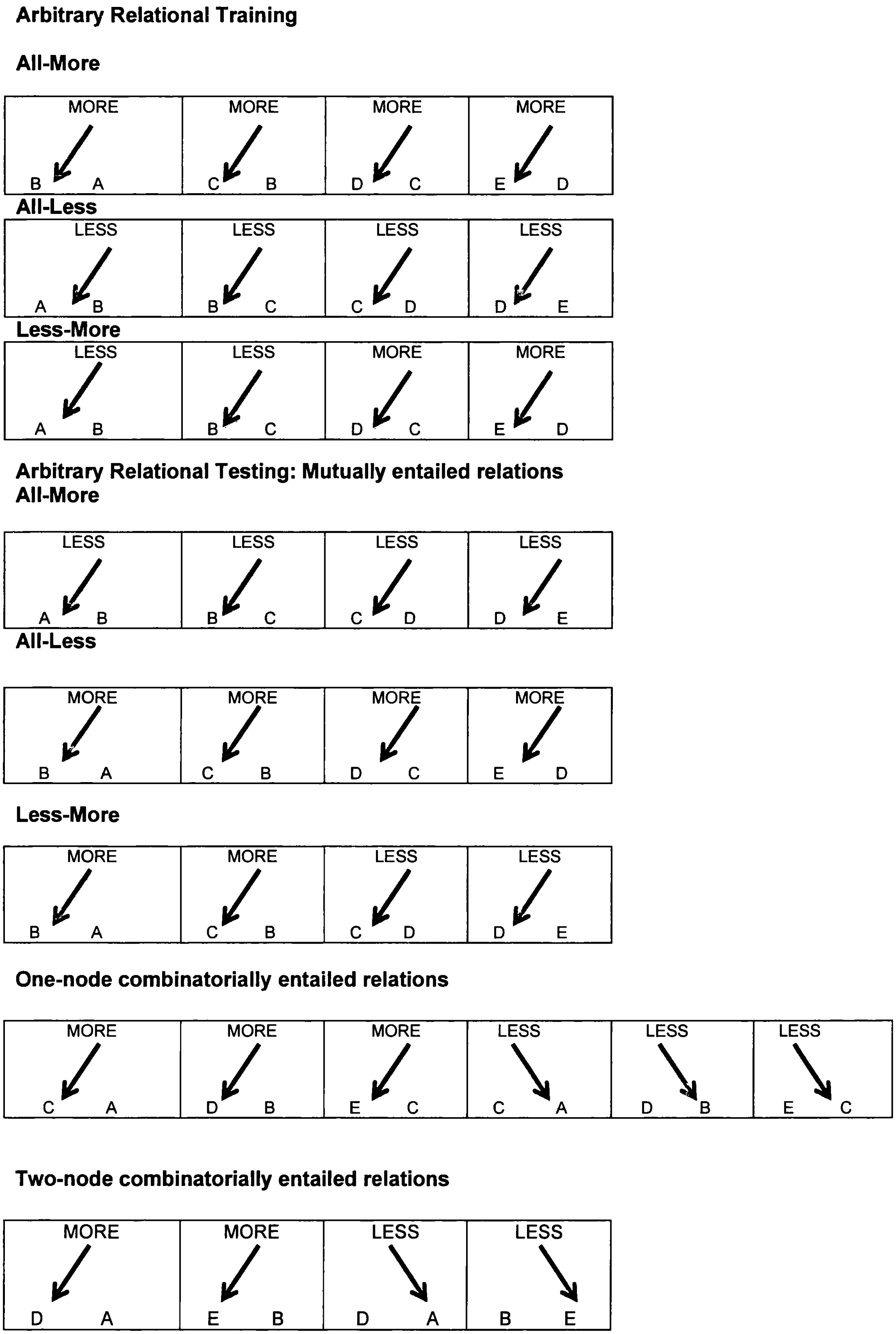


Figure 4.4. The arbitrary relational training annd test trials that participants in the All-More, All-Less and Less-More training groups were exposed to in Experiments 6A. The red arrow points to the predicted correct response.

Phase 4: Arbitrary Relational Test 1. Upon reaching arbitrary relational training criterion, participants were exposed to an arbitrary relational test phase that probed for the properties of mutual entailment alongside maintenance of the baseline arbitrary training relations. All feedback was now omitted and participants were presented with eight test trials, each presented four times, which resulted in a total of thirty-two test trials (see Figure 4.4 and Table 4.1). Dependent on the training group, the mutually entailed test trials presented to participants differed between the groups. For example, the mutually entailed relations presented to the All-More group were A<B, B<C, C<D and D<E; the All-Less group: B>A, C>B, D>C and E>D; and the Less-More group: B>A, C>B, C<D and D<E.

Table 4.1
Training and test trials received by the three groups in Experiments 6A-8B.

Group	Relation Type	Test Trial Type					
		Phase 4: Arbitrary Relational Test 1 Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
	ME	B>A	C>B	D>C	E>D		
All-More	Baseline	B>A	C>B	D>C	E>D		
	ME	A<B	B<C	C<D	D<E		
Less-More	Baseline	A<B	B<C	D>C	E>D		
	ME	B>A	C>B	C<D	D<E		
		Phase 5: Arbitrary Relational Test 2 Specific Relations in Each Group					
All-Less	Baseline	A<B	B<C	C<D	D<E		
All-More	Baseline	B>A	C>B	D>C	E>D		
Less-More	Baseline	A<B	B<C	D>C	E>D		
		Relations Common to All Groups					
	CE1	C>A	D>B	E>C	A<C	B<D	C<E
	CE2	D>A	E>B	A<D	B<E		

Note. "Baseline" refers to test trials involving directly trained relations presented in the absence of feedback; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. The inequality symbols, < (LESS-THAN) and > (MORE-THAN), denote the contextual cue presented: This indicates which comparison should be "selected over" the other, with the reinforced comparison to the left, and the unreinforced comparison to the right of the inequality symbol. It is important to note that the actual contextual cues used in the present study consisted of abstract visual images, and not the inequality symbols described here, which are used for the purposes of clarity.

Mastery criterion for this phase was set at a minimum mean of 12 out of 16 (i.e., 75% accuracy) correct responses on the baseline relations. For the mutually entailed relations, participants were required to make 3 out of 4 correct responses (i.e., 75% accuracy) on each individual mutually entailed test trial. If participants were successful in meeting criterion on both the baseline and mutually entailed relations, they progressed to a second arbitrary relational test phase. However, if participants failed to reach this mastery criterion, they were re-exposed to the experimental task from the very beginning for a maximum of three further exposures.

Phase 5: Arbitrary Relational Test 2. This test phase commenced immediately upon the successful completion of Phase 4. Participants were presented with probes for one- and two-node combinatorially entailed relations, alongside the four baseline relations. Each test trial was presented four times, in a quasi-random order, which resulted in a total of 56 test trials (see Figure 4.4 and Table 4.1). Participants were again required to make a minimum mean of 12 out of 16 correct responses on the baseline relations. All participants were also presented with same one- and two-node combinatorially entailed relations. Across these probe trials, participants were required to make a minimum of 3 out of 4 correct responses on each individual one- and two-node test trial. If participants were successful in meeting Test 2 criterion, the experiment ended and participants were asked to report to the experimenter. However, if this criterion was not met, participants were re-exposed to the entire task from Phase 1, for a maximum of three further exposures.

Results and Discussion

Of the twelve participants that took part in Experiment 6A, two (P2 and P3) failed to achieve the pre-determined Phase 4 arbitrary relational test (mutual entailment) criterion within the maximum four exposures to testing (see Table 4.2). However, ten participants successfully completed the experiment with the total number of test exposures required for these participants, ranging between 1 and 3 (M = 1.50, SD = .85). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Table 4.2 displays the number of trials to criterion for all participants during the non-arbitrary relational training, constructed-response non-arbitrary relational training and constructed-response arbitrary relational training phases. During Phase 1A (non-arbitrary relational training), participants required between 10 and 41 (M = 26.40, SD = 12.97) trials, during Phase 2A (constructed-response non-arbitrary relational training), between 10 and 17 (M = 15.90, SD = 10.50) trials, and during constructed-response arbitrary relational training, between 12 and 132 (M = 104.00, SD = 138.14) trials, to achieve criterion.

All ten participants that passed the experimental task made no errors on the baseline relations during Phase 4 (Test 1: mutual entailment). When exposed to the mutually entailed relations during Phase 4, nine participants (P1, P4, P5, P6, P7, P8, P9, P11 and P12) made no errors, while one participant (P10) made one error. When exposed to Phase 5 (Test 2: combinatorial entailment), all ten participants again made no errors on both the baseline and two-node relations. When exposed to the one-node relations, eight participants (P1, P5, P6, P7, P8, P9, P10 and P12) made no errors, while two participants (P4 and P11) made one error.

Participants who failed to pass the arbitrary relational test

As mentioned, two participants failed to achieve criterion during Phase 4 (Test 1: mutual entailment). During this test phase, P2 made between 7 and 10 out of 16 correct responses on the baseline relations, while P3 made between 6 and 10 correct responses. When exposed to the mutually entailed relations, P2 made between 7 and 10 out of 16 correct responses, while P3 made between 6 and 10 correct responses. Both participants were then excused from further participation in the experiment.

Table 4.2

Trials to criterion for the All-Less, All-More and Less-More training groups during non-arbitrary, constructed-response non-arbitrary and constructed-response arbitrary relational training in Experiment 6A. Also displayed is individual data for the All-Less, All-More and Less-More groups.

Participant	Phases 1A&2A: NARB& CR-NARB Relational Training		Phase 3: CR-Arbitrary Relational Training	Phase 4: Arbitrary Relational Test 1 (mutual entailment) B	ME	Phase 5: Arbitrary Relational Test 2 (combinatorial entailment) B	CE1	CE2
All-Less								
1	10	10	24	16/16	16/16	16/16	24/24	16/16
2*	12	10	132	8/16	11/16			
	10	10	12	8/16	12/16			
	13	10	12	7/16	8/16			
	10	10	12	10/16	10/16			
3*	22	10	132	8/16	9/16			
	15	10	24	10/16	4/16			
	10	10	60	7/16	10/16			
	10	10	12	6/16	7/16			
4	41	10	48	16/16	16/16	16/16	23/24	16/16
All-More								
5	12	10	48	3/16	0/16			
	10	10	24	16/16	15/16	14/16	16/24	14/16
	20	10	12	16/16	16/16	16/16	24/24	16/16
6	22	10	36	16/16	16/16	16/16	24/24	16/16
7	13	10	36	16/16	16/16	15/16	19/24	16/16
	13	10	12	16/16	16/16	16/16	24/24	16/16
8	15	10	60	16/16	16/16	16/16	24/24	16/16
Less-More								
9	40	10	132	16/16	16/16	16/16	24/24	16/16
10	17	10	60	16/16	15/16	16/16	24/24	16/16
11	12	10	48	16/16	16/16	16/16	23/24	16/16
12	10	17	36	16/16	15/16	16/16	14/24	16/16
	14	12	12	16/16	15/16	16/16	20/24	16/16
	15	10	12	16/16	16/16	16/16	24/24	16/16

Note. NARB = Non-arbitrary and CR = Constructed-response. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2. * refers to participants who failed to complete the experiment.

In summary, the results of Experiment 6A demonstrated that ten out of twelve participants displayed high levels of accuracy on the baseline, mutually entailed, and combinatorially entailed relations during the critical arbitrary relational test phases. In addition, there was little difference between the three groups' accuracy during testing. Thus, the RCP training and testing protocol was successful in establishing arbitrarily applicable comparative responding to "More-than" and "Less-than" relations.

Experiment 6B

In Experiment 6A, the arbitrary relational training tasks for the three training groups were presented in a linear sequence. However, as mentioned in the introduction, previous research suggests that the order in which the training pairs are presented may have an effect on test performance (Hunter, 1957; Russell et al., 1996). More specifically, Hunter (1957) argues that performances at test are weakened when training pairs are presented in a non-linear order (i.e., randomly) compared to a linear order (i.e., sequentially). In addition, Russell et al. (1996) proposed that transitive test pairs such as B<D and A<D are easier to solve (i.e., select B over D and A over D) when training pairs are presented in a linear order (e.g., A<B, B<C, C<D and D<E) compared to a non-linear order (e.g., B<C, D<E, A<B and C<D).

Experiment 6B therefore, sought to replicate and extend the findings of Experiment 1 by presenting the arbitrary relational training trials in a non-linear, sequential order.

Method

Participants

Twelve students, seven male and five female students, ranging in age from 19 to 34 years (M_{age} = 22.58, *SD* = 4.34), were recruited through student email and the psychology subject pool at Swansea University. Participants were allocated partial subject pool credit or paid £7 on completion of the study, and were randomly assigned to one of three training groups at the start of the experiment (All-More, All-Less or Less-More).

Procedure

The procedure for Experiment 6B was identical to that of Experiment 6A, with the exception of the arbitrary relational training phase (Phase 3). In this phase, the training pairs were now presented in a non-linear order. That is, in each group, the computer program quasi-randomly presented each of the four arbitrary relational training trials (Table 4.1), with the constraint that each trial could not appear more than twice consecutively. Mastery criterion for this training phase and all other phases remained the same.

Results and Discussion

Of the twelve participants that took part in Experiment 6B, one participant (P5) failed to meet criterion on the arbitrary relational test, within four exposures to testing. However, eleven participants successfully completed both arbitrary relational test phases, and required between 1 and 3 (M = 1.45, SD = .69) exposures to testing to do so (see Table 4.3). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Seven (P2, P3, P7, P8, P10, P11 and P12) of the eleven participants that passed made no errors on the baseline relations during Phase 4 (Test 1: mutual entailment). Another three participants (P1, P6 and P9) made one error, while another participant (P4) made 13 three errors. When exposed to the mutually entailed relations during Phase 4 (Test 2: combinatorial entailment), nine of the eleven participants (P2, P3, P4, P6, P8 P9, P10, P11 and P12) made no errors, while the remaining two participants (P1 and P7) made one error.

When exposed to Phase 5 (Test 2: combinatorial entailment), eight of the eleven participants (P2, P3, P4, P6, P7, P10, P11 and P12) made no errors on the baseline relations, while two participants (P8 and P9) made one error, and another participant (P1) made four errors. When exposed to the one-node relations, nine participants (P1, P2, P3, P6, P7, P8, P9, P10 and P12) made no errors, one participant (P4) made one error, and the remaining participant (P11) made three errors. Finally, during Phase 5, eight participants (P2, P4, P6, P7, P8, P9, P10 and P12) made no errors on the two-node relations, while three participants (P1, P3 and P11) made one error.

Participants who failed to pass the arbitrary relational test

P5 successfully passed tests for mutual entailment across all exposures to this test phase. P5 made between 15 and 16 out of 16 correct responses on the baseline relations during Phase 4 (Test 1: mutual entailment), and between 14 and 16 out of 16 correct responses on the mutually entailed relations during this same test phase. P5 failed to reach criterion during Phase 5 (Test 2: combinatorial entailment) across all four exposures to this test phase. P5 made between 15 and 16 out of 16 correct responses on the baseline relations, between 8 and 9 out of 24 correct responses on the one-node relations, and between 8 and 9 out of 16 correct responses on the two-node relations during tests for combinatorial entailment.

Table 4.3

Trials to criterion for the All-Less, All-More and Less-More training groups during non-arbitrary, constructed-response non-arbitrary and constructed-response arbitrary relational training in Experiment 6B. Also displayed is individual data for the All-Less, All-More and Less-More groups.

Participant	Phases 1A&2A: NARB & CR-NARB Relational Training		Phase 3: CR-Arbitrary Relational Training	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
				B	ME	B	CE1	CE2
All-Less								
1	17	10	48	15/16	15/16	12/16	24/24	15/16
2	11	10	36	16/16	16/16	16/16	24/24	16/16
3	12	10	24	16/16	16/16	16/16	24/24	15/16
4	19	10	60	13/16	16/16	16/16	23/24	16/16
All-More								
5*	10	11	108	16/16	16/16	16/16	9/24	8/16
	15	10	12	16/16	14/16	15/16	8/24	8/16
	10	10	12	16/16	16/16	16/16	8/24	9/16
	10	10	12	15/16	16/16	16/16	9/24	8/16
6	10	10	48	15/16	16/16	16/16	24/24	16/16
7	13	10	72	16/16	15/16	16/16	24/24	16/16
8	12	10	72	13/16	16/16	16/16	8/24	10/16
	10	10	12	16/16	16/16	15/16	24/24	16/16
Less-More								
9	26	10	72	15/16	16/16	15/16	24/24	16/16
10	12	10	36	15/16	16/16	16/16	8/24	0/16
	10	10	12	16/16	16/16	16/16	7/24	1/16
	10	10	12	16/16	16/16	16/16	24/24	16/16
11	15	10	84	16/16	16/16	16/16	19/24	15/16
	10	10	12	16/16	16/16	16/16	21/24	15/16
12	12	10	72	16/16	15/16	16/16	12/24	16/16
	10	10	12	16/16	16/16	16/16	24/24	16/16

Note. NARB = Non-arbitrary and CR = Constructed-response. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2. * refers to participants who failed to complete the experiment.

In summary, the results of Experiment 6B demonstrated that eleven out of twelve participants successfully passed tests for derived comparative relations when the arbitrary relational training tasks were presented in a non-linear order. There was little difference between the three training groups, with all groups displaying high levels of accuracy on the baseline, mutually entailed, and one- and two-node combinatorially entailed relations. Thus, the RCP was successful in establishing responding in accordance with the relational frames of "More-than" and "Less-than" when the arbitrary relations were trained in a non-linear sequence.

The findings from Experiments 6A and 6B taken together appear to have overcome some of the problems observed with the lack of stimulus control in Chapter 3. For example, as can be seen from the results described for individual participants in both Experiments 6A and 6B, responding to the mutually entailed relations improved considerably over those reported in Chapter 3. For participants that passed Experiments 6A and 6B, high accuracy was observed on the mutually entailed relations (Experiment 6A: P1 and P's 4-12; Experiment 6B: P's 1-4and P's 6-12). Similarly, high accuracy was observed for these participants on the one- and two-node relations. In contrast, across Experiments 4-5B in Chapter 3, a number of participants failed to make any correct responses to the mutually entailed relations, and thus, it was proposed, that participants were responding to test relations in a similar manner to training. Indeed, when interventions by means of an observing response and a variant of the simple-to-complex protocol were introduced in Experiments 5A and 5B in Chapter 3, only limited improvements in stimulus control were noted. On the other hand, results from Experiments 6A and 6B of Chapter 4 show that the failures in stimulus control were greatly reduced. Thus, incorporating a variant of the RCP to establish derived comparative responding appears to have overcome some of the problems encountered in Chapter 3. However, it must also be noted that responding for two participants in Experiment 6A (P2 and P3) and one participant in 6B (P5) did not come under effective stimulus control. Indeed, for participants in Experiment 6A that failed to meet criterion during arbitrary relational testing (P2 and P3), both were unable to respond in accordance with the properties of mutual entailment. Future studies should therefore seek to include appropriate training interventions (e.g., non-arbitrary relational training) in an attempt to generate more accurate performances.

Experiment 7A

As a result of the large number of participants successfully completing Experiments 6A and 6B, it was questioned as to whether the non-arbitrary relational training and testing phases, in addition to the constructed-response non-arbitrary relational training and testing phases, were necessary to generate derived comparative responding. For example, identifying efficient methods of generating such patterns of responding is important as the minimum time in which participants could cycle through the RCP protocol in Experiments 6A and 6B was forty minutes. Furthermore, if participants did not meet test criterion, they were re-exposed to the experimental session from the very beginning, which resulted in a session lasting between one and two hours. Thus, the development of efficient training and testing protocols is an important issue in studies examining derived stimulus relations. In Experiments 7A and 7B therefore, the first non-arbitrary relational training and testing phase was omitted, and participants were exposed to the experimental task from the constructed-response non-arbitrary relational training and testing phases. In addition, participants in Experiment 7A, were exposed to the arbitrary baseline relations in a linear order, while participants in Experiment 7B were exposed to the arbitrary baseline relations in a non-linear order.

Method

Participants

Twelve students, one male and eleven female, ranging in age from 18 to 28 years (M_{age} = 20.17, SD = 2.62) were recruited through student email and the psychology subject pool at Swansea University. Participants were allocated partial subject pool credit or paid £7 on completion of the study, and were randomly assigned to one of three training groups at the start of the experiment (All-More, All-Less or Less-More).

Procedure

The general procedure of Experiment 7A was similar to that of Experiments 6A and 6B. However, in this experiment, the first phase involving non-arbitrary relational training and testing was now omitted. Therefore, participants were exposed to *Phase*

1A: Constructed-Response Non-arbitrary Relational Training and *Phase 1B: Constructed-Response Non-arbitrary Relational Testing; Phase 2: Constructed-Response Arbitrary Relational Training; Phase 3: Arbitrary Relational Test 1 (mutual entailment)* and *Phase 4: Arbitrary Relational Test 2 (combinatorial entailment).*

Phases 1A and 1B: Constructed-Response Non-Arbitrary Relational Training and Testing. Both phases were identical to Phases 2A and 2B of Experiments 6A and 6B.

Phase 2: Constructed-Response Arbitrary Relational Training. This phase was identical to Phase 3 in Experiment 6A, where all training pairs were presented in a linear order.

Phases 3 and 4: Arbitrary Relational Test 1 (mutual entailment) and Arbitrary Relational Test 2 (combinatorial entailment). Both phases were identical to Phases 4 and 5 respectively, in Experiments 6A and 6B.

Mastery criterion for all training and testing phases remained the same as in Experiments 6A and 6B.

An additional feature of this experiment was that participants were now required to make their responses on a touch-screen computer monitor. Therefore, instead of using the computer mouse to make their responses, participants were required to touch the comparison stimulus they had selected and drag it across the computer monitor to the blank yellow squares. This response requirement was in force during all training and testing phases in Experiment 7A.

Results and Discussion

Of the twelve participants that began Experiment 7A, ten participants failed to successfully complete the experimental task within the pre-determined four exposures to testing. Only two participants (P1 and P5) successfully completed the experimental task, with both participants doing so on their first exposure to Tests 1 and 2 (see Table 4.4). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Table 4.4 displays the trials to criterion for participants during the constructed-response non-arbitrary relational training, and constructed-response arbitrary relational training phases. During Phase 1A (constructed-response non-arbitrary relational training), participants took between 10 and 16 (M = 12.00, SD = .00) trials to achieve training criterion. Participants took between 12 and 108 (M = 36.00, SD = 16.92) trials across all exposures to training, to achieve mastery criterion for the constructed-response arbitrary relational training phase.

For the two participants that successfully completed the experimental task, P1 made no errors on the baseline and mutually entailed relations during Phase 3 (Test 1: mutual entailment; see Table 4.4), while P5 made one error on the baseline relations and no errors on the mutually entailed relations. When exposed to Phase 4 (Test 2: combinatorial entailment), both participants made no errors on the baseline, and one- and two-node relations.

Participants who failed to pass the arbitrary relational test

Five participants (P6, P9, P10, P11 and P12) failed to achieve the pre-determined mastery criterion during Phase 3 (Test 1: mutual entailment). Across four exposures to this test phase, P6 made between 9 and 15 out of 16 correct responses on the baseline relations, and between 9 and 14 out of 16 correct responses on the mutually entailed relations. On the same four exposures, P9 made between 9 and 13 correct responses on the baseline relations, and between 5 and 12 correct responses on the mutually entailed relations. P10 made no errors on the baseline relations, but made 0 out of 16 correct responses on the mutually entailed relations, across four exposures to Phase 3. P11 made between 15 and 16 correct responses on the baseline relations, but only 4 out of 16 correct responses on the mutually entailed relations, across four exposures to this test phase. Finally, P12 made between 8 and 16 correct responses on the baseline relations, and between 1 and 12 correct responses on the mutually entailed relations, across his/her four exposures to Phase 3.

Table 4.4
Trials to criterion for the All-Less, All-More and Less-More training groups during constructed-response non-arbitrary, and constructed-response arbitrary relational training in Experiment 7A. Also displayed is individual data for the All-Less, All-More and Less-More groups.

	Phase 1A: CR-NARB Relational	Phase 2: CR-Arbitrary Relational	Phase 3: Arbitrary Relational Test 1 (mutual entailment)		Phase 4: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training	Training	B	ME	B	CE1	CE2
All-Less							
1	12	24	16/16	16/16	16/16	24/24	16/16
2*	12	48	15/16	16/16	12/16	7/24	9/16
	10	24	16/16	16/16	16/16	12/24	12/16
	10	12	16/16	16/16	15/16	16/24	16/16
	10	12	16/16	16/16	16/16	16/24	16/16
3*	12	96	15/16	15/16	16/16	16/24	14/16
	10	12	15/16	16/16	12/16	15/24	15/16
	10	12	16/16	15/16	13/16	16/24	13/16
	10	24	16/16	14/16	16/16	15/24	16/16
4*	12	60	10/16	9/16			
	10	12	16/16	16/16	13/16	12/24	10/16
	12	12	15/16	15/16	14/16	11/24	11/16
	10	12	12/16	12/16			
All-More							
5	12	48	16/16	15/16	16/16	24/24	16/16
6*	12	72	15/16	9/16			
	10	12	11/16	10/16			
	10	12	9/16	11/16			
	10	24	13/16	14/16			
7*	12	108	13/16	16/16	11/16	13/24	7/16
	10	24	16/16	16/16	16/16	9/24	9/16
	10	12	16/16	16/16	16/16	8/24	16/16
	10	12	15/16	16/16	16/16	8/24	15/16
8*	12	96	16/16	15/16	16/16	13/24	4/16
	10	12	16/16	16/16	16/16	9/24	0/16
	12	12	16/16	4/16			
	10	12	16/16	0/16			
Less-More							
9*	12	36	13/16	6/16			
	10	24	12/16	5/16			
	10	24	10/16	12/16			
	10	24	9/16	5/16			
10*	16	24	16/16	0/16			
	10	12	16/16	0/16			
	10	12	16/16	0/16			
11*	10	12	16/16	0/16			
	21	36	15/16	4/16			
	16	24	16/16	4/16			

Table 4.4 (*cont.d*)	Phase 1A: CR-NARB Relational	Phase 2: CR-Arbitrary Relational	Phase 3: Arbitrary Relational Test 1 (mutual entailment)		Phase 4: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training	Training	B	ME	B	CE1	CE2
11* (*cont.d*)	10	12	16/16	4/16			
12*	12	60	16/16	3/16			
	11	12	11/16	1/16			
	10	12	8/16	7/16			
	10	12	15/16	12/16			

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2. * refers to participants who failed to complete the experiment.

For the remaining five participants that took part in Experiment 7A, three participants (P2, P3 and P7) were exposed to both Phase 3 (Test 1: mutual entailment) and Phase 4 (Test 2: combinatorial entailment). Across four exposures to both test phases, P2 made between 15 and 16 out of 16 correct responses on the baseline relations, and 16 out of 16 correct responses on the mutually entailed relations during Phase 3. When exposed to Phase 4, P2 made between 12 and 16 out of 16 correct responses on the baseline relations, between 7 and 16 out of 24 correct responses on the one-node relations, and between 9 and 16 out of 16 correct responses on the two-node relations. P3, across four exposures to both test phases, made between 15 and 16 correct responses on the baseline relations, and between 14 and 16 correct responses on the mutually entailed relations, when exposed to Phase 3. When exposed to Phase 4, P3 made between 12 and 16 correct responses on the baseline relations, between 15 and 16 correct responses on the one-node relations, and between 13 and 16 correct responses on the two-node relations. Across four exposures to both test phases, P7 made between 13 and 16 correct responses on the baseline relations, and 16 out of 16 correct responses on the mutually entailed relations during Phase 3. When exposed to Phase 4, P7 made between 11 and 16 correct responses on the baseline relations, between 8 and 13 correct responses on the one-node relations and between 7 and 16 correct responses on the two-node relations.

For the final two participants (P4 and P8), both were exposed to Phase 3 for a total of four times and Phase 4 for a total of two times. Across four exposures to Phase 3, P4 made between 10 and 16 correct responses on the baseline relations and between 9 and 16 correct responses on the mutually entailed relations. Across two exposures to Phase 4, P4 made between 13 and 14 correct responses on the baseline relations, between 1 and 12 correct responses on the one-node relations and between 10 and 11 correct responses on the two-node relations. P8, across four exposures to Phase 3, made no errors on the baseline relations, and made between 0 and 16 correct responses on the mutually entailed relations. On his/her two exposures to Phase 4, P8 again made no errors on the baseline relations, and between 9 and 13 correct responses on the one-node relations, and between 0 and 4 correct responses on the two-node relations.

In summary, only two out of the twelve participants (P1 and P5) that started Experiment 7A managed to successfully complete the experimental task. Five participants were unable to pass tests for mutual entailment (P6, P9, P10, P11 and P12), while three particopants (P2, P3 and P7) encountered difficulties on the two-node relations, and two participants (P4 and P8) encountered difficulties on the mutually entailed and one- and two-node relations. Thus, the findings from Experiment 7A failed to replicate those of Experiments 6A and 6B. These findings are similar to those from Experiments 4-5B in Chapter 3 in that, for a number of participants, stimulus control was largely absent. More specifically, and again similar to findings from Chapter 3, the contextual cues for MORE-THAN and LESS-THAN did not appear to be functioning during the arbitrary relational phases. Thus, in Experiment 7A, the variant of the RCP did not lead to improvements in stimulus control over the conditional discrimination, employed in Chapter 3. However, one potential reason for the observed findings may be due to the reduced number of non-arbitrary training trials that participants were exposed to. For instance, participants in Experiments 6A and 6B were exposed to an additional non-arbitrary relational training and test phase, over those in Experiment 7A. Therefore, it may be possible that the reduced number of training trials were not sufficient to establish effective stimulus, or contextual control in Experiment 7A. Future studies should seek to determine the number of non-arbitrary training trials required to establish effective stimulus control.

Experiment 7B

Despite the fact that a large number of participants failed to successfully complete Experiment 7A, Experiment 7B again examined the emergence of arbitrary comparative performances using the RCP. Similar to Experiment 7A, participants were not exposed to the first non-arbitrary relational training and test phase, and instead, were exposed to the experimental task from constructed-response non-arbitrary relational training. In addition, Experiment 7B also sought to further our understanding of the effects of linearity on relational performances at test. Thus, participants in Experiment 7B were exposed to the training pairs in a non-linear order.

Method

Participants

Twelve participants, three male and nine female, ranging in age from 18 to 27 years (M_{age} = 21.58, *SD* = 2.64) were recruited through student email and the psychology subject pool at Swansea University. Participants were allocated partial subject pool credit or paid £7 on completion of the study, and were randomly assigned to one of three training groups at the start of the experiment (All-More, All-Less or Less-More).

Procedure

The procedure for Experiment 7B was identical to that of Experiment 7A, with the exception that during the constructed-response arbitrary relational training phase, pairs were now presented in a non-linear order. Mastery criterion for this training phase and all other phases remained the same.

Results and Discussion

Of the twelve participants that started Experiment 7B, seven participants (P1, P2, P3, P6, P7, P10 and P11) successfully completed the experimental task (see Table 4.5). The remaining five participants (P4, P5, P8 P9 and P12) failed to successfully complete the experimental task within the maximum four exposures to testing. The number of test exposures for participants that passed ranged between 1 and 4 (*M* =

1.63, *SD* = .74). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Table 4.5 displays the trials to criterion for the constructed-response non-arbitrary relational training phase, and the constructed-response arbitrary relational training phase. Participants took between 10 and 24 (*M* = 27.00, *SD* = 8.45) trials to complete constructed-response non-arbitrary relational training and between 12 and 144 (*M* = 94.29, *SD* = 54.30) trials to complete constructed-response arbitrary relational training, across all four exposures to the experimental task.

Of the seven participants that passed, five (P3, P6, P7, P10 and P11) made no errors on the baseline relations during Phase 3 (Test 1: mutual entailment), while two participants (P1 and P2) made one error. When exposed to the mutually entailed relations, six participants (P1, P2, P3, P6, P7 and P11) made no errors, while one participant (P10) made one error.

When exposed to Phase 4 (Test 2: combinatorial entailment), six participants (P1, P3, P6, P7, P10 and P11) made no errors on the baseline relations, while one participant (P2) made four errors. When exposed to the one-node relations, five participants (P1, P3, P6, P7 and P10) made no errors, while one participant (P2) made one error, and another participant (P11) made two errors. Finally, when exposed to the two-node relations, all seven participants made no errors on these relations.

Table 4.5

Trials to criterion for the All-Less, All-More and Less-More training groups during constructed-response non-arbitrary and constructed-response arbitrary relational training in Experiment 7B. Also displayed is individual data for the All-Less, All-More and Less-More groups.

	Phase 1A: CR-NARB Relational	Phase 2: CR-Arbitrary Relational	Phase 3: Arbitrary Relational Test 1 (mutual entailment)		Phase 4: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training	Training	B	ME	B	CE1	CE2
All-Less							
1	21	96	15/16	16/16	16/16	24/24	16/16
2	13	144	15/16	16/16	13/16	21/24	15/16
	10	12	15/16	12/16			
	10	24	15/16	16/16	12/16	23/24	16/16
3	16	84	16/16	14/16	15/16	20/24	13/16
	10	24	16/16	16/16	16/16	24/24	16/16
4*	12	60	12/16	6/16			
	10	24	14/16	13/16			
	10	12	11/16	13/16			
	10	12	14/16	14/16			
All-More							
5*	13	60	14/16	12/16			
	10	24	16/16	16/16	16/16	9/24	1/16
	10	12	16/16	15/16	16/16	10/24	0/16
	10	12	15/16	12/16			
6	19	48	16/16	16/16	16/16	24/24	16/16
7	12	36	16/16	16/16	16/16	24/24	16/16
8*	12	60	14/16	13/16			
	10	24	16/16	15/16	16/16	17/24	16/16
	10	12	16/16	16/16	16/16	17/24	16/16
	10	12	16/16	16/16	16/16	16/24	16/16
Less-More							
9*	12	48	16/16	0/16			
	10	12	16/16	0/16			
	10	12	16/16	0/16			
	10	12	16/16	0/16			
10	12	144	16/16	15/16	16/16	24/24	16/16
11	21	36	15/16	5/16			
	10	12	16/16	16/16	16/16	22/24	16/16
12*	12	36	11/16	4/16			
	15	36	10/16	11/16			
	10	12	16/16	12/16			
	18	12	16/16	10/16			

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2. * refers to participants who failed to complete the experiment.

Participants who failed to pass the arbitrary relational test

For the five participants that failed to successfully complete the experimental task, three participants (P4, P9 and P12) failed to reach Phase 3 (Test 1: mutual entailment) criterion across four exposures to this test phase. P4 made between 11 and 14 out of 16 correct responses on the baseline relations, and between 6 and 14 out of 16 correct responses on the mutually entailed relations, across four exposures to this test phase. P9 and P12 also failed to achieve Test 1 criterion. P9 made no errors on the baseline relations, and no correct responses on the mutually entailed relations across four exposures to testing. In addition, P12 made between 10 and 16 correct responses on the baseline relations, and between 12 and 16 correct responses on the mutually entailed relations, across four exposures to this test phase.

For the remaining two participants (P5 and P8), both were exposed to Phase 3 (Test 1: mutual entailment) for the maximum four times. P5 made between 14 and 15 out of 16 correct responses on the baseline relations and between 12 and 16 out of 16 correct responses on the mutually entailed relations across all four exposures to this test phase. During his/her two exposures to Phase 4 (Test 2: combinatorial entailment), P5 made 16 out of 16 correct responses on the baseline relations, between 9 and 10 out of 24 correct responses on the one-node relations, and between 0 and 1 out of 16 correct responses on the two-node relations. Finally, across all four exposures to Phase 3, P8 made between 14 and 16 correct responses on the baseline relations and between 14 and 16 correct responses on the mutually entailed relations. Across his/her three exposures to Phase 4, P8 made no errors on the baseline and two-node relations, and between 16 and 17 correct responses on the one-node relations.

In summary, a total of seven out of twelve participants (P1, P2, P3, P6, P7, P10 and P11) managed to successfully complete the experimental task, displaying high levels of accuracy on all test relations. Although the number of participants that passed was higher than Experiment 7A, a number of participants again failed to pass tests for mutual and combinatorial entailment. For example, three participants (P4, P9 and P12) encountered difficulties on the mutually entailed relations, while P8 was unable to accurately respond to the one-node relations. In addition, P5 encountered difficulties on the one- and two-node relations. Thus, similar to previous experiments (Experiments 4-5B and Experiment 7A), effective stimulus control over responding was absent for a number of participants. However, similar to suggestions in the discussion of findings from Experiment 7A, the reduced number of non-arbitrary training trials that participants in Experiments 7A and 7B were exposed to, may have contributed to these findings. Experiments 8A and 8B will therefore seek to explore the utility of a multiple-exemplar training (MET) intervention consisting of additional non-arbitrary training trials in an attempt to establish greater stimulus control over responding.

Experiment 8A

Experiments 8A and 8B were designed in an attempt to address the findings from Experiments 7A and 7B that the predicted derived arbitrary comparative relational performances were found to be lacking, or absent. Both experiments investigated the potential utility of non-arbitrary trials, as a form of MET, in facilitating this type of responding. For example, a number of studies have reported the utility of MET in the establishment or facilitation of derived relational responding in both adults and children (e.g., Barnes-Holmes et al., 2004; Berens & Hayes, 2007; Gorham et al., 2009; Luciano, Gomez Becerra, & Rodriguez Valverde, 2007; Vitale et al., 2008). In one such study, Berens and Hayes (2007) exposed four children to a MET intervention with non-arbitrary stimulus sets when it was found that arbitrary comparative responding was deficient. This intervention consisted of placing piles of pennies on the picture cards that were used to train the arbitrary relations. For example, on a trial in which a "More-than" relation was being trained, there were more pennies on the picture card that was specified as "more", in comparison to the picture card specified as "less". The authors found that the intervention was successful in facilitating the development of arbitrary comparative relations, and that these abilities generalised to other trial types. In addition, Vitale et al. (2008) noted improvements in weak relational performances at test when participants were exposed to arbitrary test problems in non-arbitrary form. Therefore, participants in Experiments 8A and 8B were first exposed to constructed-response non-arbitrary relational training and testing, followed by a MET intervention that consisted of training and testing with novel non-arbitrary stimulus sets.

Lastly, Experiments 8A and 8B again sought to explore the effects of linearity on these performances, and participants in Experiment 8A were exposed to the arbitrary baseline relations in a linear order.

Method

Participants

Four female postgraduate students, ranging in age from 24 to 30 years (M_{age} = 27.50, *SD* = 3.05) were recruited through personal contacts. Participants received no reimbursement for their participation in the study, and were randomly assigned to one of three training groups at the start of the experiment (All-More, All-Less or Less-More).

Procedure

The general procedure was identical to Experiments 7A and 7B. However, an additional constructed-response non-arbitrary relational training and testing phase was included as a multiple-exemplar training intervention, following Phases 1A and 1B. Both phases were identical to Phases 1A and 1B, with the exception that participants were exposed to novel training and testing stimulus sets. In addition, during constructed-response arbitrary relational training, the training pairs were presented in a linear order. All other phases and criterion remained the same as Experiments 7A and 7B.

Results and Discussion

Of the four participants that took part in Experiment 8A, all four successfully completed the experimental task (see Table 4.6), with the number of exposures to testing for participants ranging between 1 and 3 (*M* = 1.50, *SD* = 1.00). Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Table 4.6 displays the trials to criterion for participants during both constructed-response non-arbitrary relational training phases and constructed-response arbitrary relational training. All four participants required between 10 and 16 (Phase 1A: M = 18.00, SD = 12.33; Phase 2A: M = 12.00, SD = 10.00) trials to reach criterion during the constructed-response non-arbitrary relational training phases. The number of trials required to achieve constructed-response arbitrary relational training ranged between 12 and 48 (M = 36.00, SD = 21.91).

When exposed to Phase 4 (Test 1: mutual entailment), three participants (P2, P3 and P4) made no errors on the baseline relations, while one participant (P1) made one error. When exposed to the mutually entailed relations, all four participants made no errors on these relations.

When exposed to Phase 5 (Test 2: combinatorial entailment), three participants (P1, P3 and P4) made no errors on the baseline relations, while one participant (P2) made one error. When exposed to the one-node relations, two participants (P1 and P4) made no errors, while one participant (P3) made one error, and another (P2) made three errors. Finally, when exposed to the two-node relations, three participants (P1, P2 and P3) made no errors, while one participant (P4) made one error.

Table 4.6
Trials to criterion for the All-Less, All-More and Less-More training groups during constructed-response non-arbitrary and constructed-response arbitrary relational training in Experiment 8A. Also displayed is individual data for the All-Less, All-More and Less-More groups.

	Phases 1A&2A: CR-NARB Relational		Phase 3: CR-Arbitrary Relational	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training		Training	B	ME	B	CE1	CE2
1(All-Less)	16	10	48	15/16	16/16	16/16	24/24	16/16
2(All-Less)	16	10	24	11/16	12/16			
	10	10	12	15/16	15/16	16/16	22/24	10/16
	10	10	24	16/16	16/16	15/16	21/24	16/16
3(All-More)	10	10	12	16/16	16/16	16/16	23/24	16/16
4(Less-More)	10	10	24	16/16	16/16	16/16	24/24	15/16

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2.

In summary, all four participants in Experiment 8A successfully completed the experimental task, and displayed high levels of accuracy on all test relations. Thus, the increased number of non-arbitrary trials that participants in Experiment 8A were exposed to, appear to have overcome the failures in stimulus control encountered in Experiments 7A and 7B. However, the impact of the non-arbitrary training intervention on arbitrary relational performances at test remains unclear. As all participants were exposed to the intervention from the beginning of the experiment, it is difficult to determine whether or not the intervention was conducive in facilitating responding to arbitrary comparative relations. In previous studies (e.g., Berens & Hayes, 2007; Gorham et al., 2009; Vitale et al., 2008), additional non-arbitrary trials were only introduced, if, during the initial probe trials, participants failed to demonstrate derived comparative responding. Therefore, future studies should seek to include additional non-arbitrary trials if these relations do not emerge immediately. This, in turn, may help to identify the conditions both necessary and sufficient for the emergence of arbitrary comparative responding.

Experiment 8B

Experiment 8B was again concerned with examining the effects of a MET intervention on arbitrary comparative performances at test. Participants were exposed to the intervention following constructed-response non-arbitrary relational training and testing. In addition, participants were also exposed to the arbitrary baseline pairs in a non-linear order.

Method

Participants

Four participants, one male and three female, ranging in age from 19 to 25 years (M_{age} = 21.00, SD = 2.71), were recruited through the psychology subject pool at Swansea University. Participants were allocated partial subject pool credit on completion of the study, and were randomly assigned to one of the three training groups at the start of the experiment (All-More, All-Less or Less-More).

Procedure

The general procedure was identical to Experiment 8A except for the following important difference: during constructed-response arbitrary relational training, pairs of comparison stimuli were presented in a non-linear order. Training and test criteria for all phases remained the same.

Results and Discussion

Four participants took part in Experiment 8B, with two participants (P1 and P3), successfully completing the experimental task on their first exposure to testing (see Table 4.7). The remaining two participants (P2 and P4) failed to achieve test mastery criterion, and thus, failed to successfully complete the experimental task. Results are discussed for participants that passed (met criterion at testing) and failed the experiment.

Table 4.7 displays the trials to criterion for all four participants during the constructed-response non-arbitrary relational training phases and the constructed-response arbitrary relational training phase. Across all exposures to testing, participants required between 10 and 20 (Phase 1A: M = 11.00, SD = 1.41; Phase 2A: M = 10.00, SD = .00) trials to achieve criterion during both constructed-response non-arbitrary relational training phases. During constructed-response arbitrary relational training, participants required between 24 and 120 (M = 36.00, SD = 16.97) trials to reach training criterion.

When exposed to Phase 4 (Test 1: mutual entailment), P1 made two errors on the baseline relations, while P3 made no errors. P1 then made one error on the mutually entailed relations, while P3 made no errors. When exposed to Phase 5 (Test 2: combinatorial entailment), both P1 and P3 made no errors on the baseline and two-node relations. When exposed to the one-node relations, P1 made one error, while P3 made two errors.

Participants who failed to pass the arbitrary relational test

Across all four exposures to Phase 4, P2 made between 13 and 15 out of 16 correct responses on the baseline relations, and between 12 and 16 out of 16 correct responses on the mutually entailed relations. P2 was exposed to Phase 5 for a total of two times. Across both exposures, P2 made between 12 and 14 out of 16 correct responses on the baseline relations, between 18 and 19 out of 24 correct responses on the one-node relations, and between 6 and 15 out of 16 correct responses on the two-node relations. Finally, across all four exposures to Phase 4, P4 made no errors on the baseline relations, and between 0 and 16 correct responses on the mutually entailed relations. On his/her one and only exposure to Phase 5, P4 made 11 correct responses on the baseline relations, 10 correct responses on the one-node relations, and 7 correct responses on the two-node relations.

Table 4.7
Trials to criterion for the All-Less, All-More and Less-More training groups during constructed-response non-arbitrary and constructed-response arbitrary relational training in Experiment 8B. Also displayed is individual data for the All-Less, All-More and Less-More groups.

	Phases 1A&2A: CR-NARB Relational		Phase 3: CR-Arbitrary Relational	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training		Training	B	ME	B	CE1	CE2
1(All-Less)	12	10	48	14/16	15/16	16/16	23/24	16/16
2(All-More)*	16	10	96	13/16	12/16			
	10	10	36	16/16	16/16	14/16	18/24	15/16
		10		15/16	15/16	12/16	19/24	6/16
	10	10	24	15/16	14/16			
3(All-More)	10	10	24	16/16	16/16	16/16	22/24	16/16
4(Less-More)*	11	10	36	16/16	16/16	11/16	10/24	7/16
	12	10	120	16/16	0/16			
	10	20	24	16/16	0/16			
	10	10	24	16/16	9/16			

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2. * refers to participants who failed to complete the experiment.

In summary, two out of four participants (P1 and P3) successfully completed the experimental task on their first exposure to testing. The remaining two participants (P2 and P4) failed to meet test criterion. For participants that failed, P2 encountered difficulties on the mutually entailed and one- and two-node relations, while P4 failed tests for mutual entailment. In contrast, in Experiment 8A, all participants met criterion at test and displayed accurate responding on all trials at test. Thus, an issue arising from the findings of Experiments 8A and 8B is whether the non-arbitrary training intervention, or the presentation of training trials in a linear order, led to the superior performances noted for participants in Experiment 8A. Results from Experiments 6A and 6B demonstrated that linearity was not found to affect arbitrary comparative performances at test, while results from Experiments 7A and 7B proposed that additional exposure to non-arbitrary training trials may be necessary to generate greater stimulus control. Further research is therefore warranted, to determine the number of non-arbitrary training trials needed to generate successful arbitrary comparative responding.

Summary of Experiments 6A-8B

Table 4.8
A summary of the relational networks, training and testing protocols, linearity of training pairs and overall yield in Experiments 6A-8B.

Experiment	Relational Network	Training & Testing Protocol	Linearity of training pairs	Yield
6A	5-term	NARB training & testing & CR	Linear	10/12
6B	5-term	NARB training & testing & CR	Non-linear	11/12
7A	5-term	CR	Linear	2/12
7B	5-term	CR	Non-linear	7/12
8A	5-term	CR	Linear	4/4
8B	5-term	CR	Non-linear	2/4

Note. The acronyms NARB and CR refer to non-arbitrary and constructed-response, respectively.

Further Analyses

Additional analyses were conducted across all experiments to examine the overall yield (i.e., the number of participants that successfully passed both arbitrary relational tests), and number of exposures to the experimental task that participants required. The analyses also examined the effectiveness of two arbitrary training designs (i.e., linear and non-linear) in facilitating the emergence of derived comparative relations.

Experiments 6A and 6B. In Experiments 6A (RCP-linear) and 6B (RCP-non-linear), there were yields of 83% (10 out of 12 participants) and 91.6% (11 out of 12 participants), respectively (see Figure 4.5). There was little difference in the overall mean number of test exposures required, between Experiments 6A (M = 1.50, SD = .85) and 6B (M = 1.45, SD = .69).

Experiments 6A and 6B demonstrated that the RCP was effective in establishing responding in accordance with the relational frames of "More-than" and "Less-than" in adult populations. The RCP with training pairs presented in a non-linear order was found to have a slight advantage over the RCP with training pairs presented in a linear order, in terms of the overall yield. This difference was not statistically significant, Mann-Whitney U, z = -.60, p = .55.

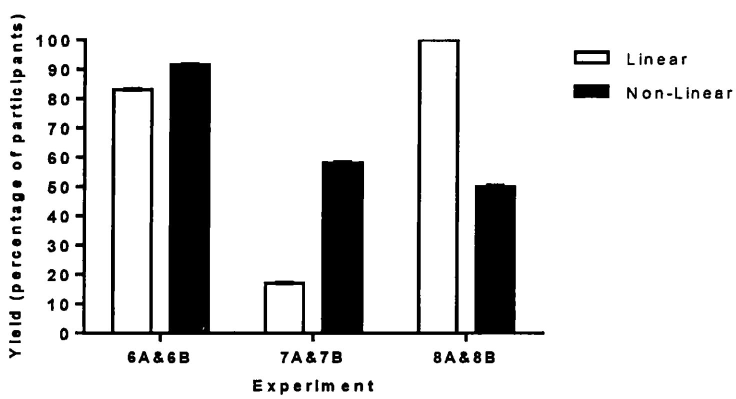


Figure 4.5. The percentage yield is displayed for Experiments 6A, 6B, 7A, 7B, 8A and 8B.

Arbitrary Training. The mean number of trials required by participants in Experiment 6A to complete arbitrary relational training was 83, while participants in Experiment 6B required, on average, 97 trials. However, this difference was non-significant, $t(22)$ = .26, p = .80.

Experiments 7A and 7B. In Experiments 7A (RCP (constructed response-linear)) and 7B (RCP (constructed response-non-linear)), there were yields of only 17% (2 out of 12 participants) and 58% (7 out of 12 participants), respectively (see

Figure 4.5). Participants that passed in Experiment 7A required fewer exposures (M = 1.00, SD = .00) to testing than participants in Experiment 7B (M = 1.57, SD = .79).

Experiments 7A and 7B found that only two participants successfully completed the experiment when the training pairs were presented in a linear order. In contrast, seven participants passed the experiment when the training pairs were presented in a non-linear order. This difference was statistically significant, Mann-Whitney U, z = -2.06, p = .04.

Arbitrary Training. The mean number of trials required by participants in Experiment 7A to complete arbitrary relational training was 69, while participants in Experiment 7B required, on average, 95 trials. However, this difference was non-significant, $t(22) = -1.84$, p = .09.

Experiments 8A and 8B. In Experiments 8A and 8B, there were yields of 100% (4 out of 4 participants) and 50% (2 out of 4 participants), respectively (see Figure 4.5). Participants in Experiment 8A (M = 1.50, SD = 1.00) required, on average, more exposures to testing than participants in Experiment 8B (M = 1.00, SD = .00).

Experiments 8A and 8B found that the RCP with training pairs presented in a linear order was successful for all four participants in establishing responding in accordance with the relational frames of “More-than” and “Less-than”. The RCP with training pairs presented in a non-linear order resulted in two out of four participants successfully completing the experiment. This difference however, was not statistically significant, Mann-Whitney U, z = -1.53, p = .13.

Arbitrary Training. The mean number of trials required by participants in Experiment 8A to complete arbitrary relational training was 36, while participants in Experiment 8B required, on average, 114 trials. However, this difference was non-significant, $t(6) = -1.77$, p = .13.

General Discussion

The current series of experiments undertook an investigation of the potential utility of the RCP and constructed-response protocol in establishing derived comparative relations. In addition, all studies examined the effects of linearity on arbitrary "More-than" and "Less-than" performances at test. Furthermore, Experiments 8A and 8B also explored the potential facilitative effects of a multiple-exemplar training intervention on the emergence of these relations.

In Experiment 6A, ten out of twelve (83% yield) participants successfully completed the experimental task, while eleven out of twelve (91.6%) participants in Experiment 6B passed. Participants were assigned to one of three training groups; All-More, All-Less and Less-More, where the training pairs differed between the groups. However, all participants were exposed to the same test relations. For the participants that passed, little difference was observed between the groups in terms of accuracy on the mutually and combinatorially entailed test relations. Thus, findings from Experiments 6A and 6B highlight the potential utility of the RCP and constructed-response protocol in generating arbitrary comparative responding.

Experiments 7A and 7B were similar to Experiments 6A and 6B, with the exception that the first phase of the experiment (non-arbitrary relational training and testing) was omitted. Only two out of twelve (17% yield) participants in Experiment 7A passed the experimental task, while seven out of twelve (58% yield) participants successfully completed Experiment 7B. For the participants that passed, accuracy was high on all test relations. Despite this, a large number of participants failed to demonstrate the emergence of derived comparative responding.

Experiments 8A and 8B were specifically designed in an attempt to explore effects of a multiple-exemplar training (MET) intervention consisting of non-arbitrary stimulus sets on arbitrary comparative responses at test. Participants were first exposed to constructed-response non-arbitrary relational training and testing, followed by the multiple-exemplar training intervention. All four (100% yield) participants in Experiment 8A passed the experimental task, while two out of four (50% yield) participants in Experiment 8B completed the task. High levels of accuracy were observed for the participants that passed on all test relations.

Improvements in stimulus control

Findings from a number of experiments (6A, 6B and 8A) in the current chapter appear to have overcome some of the failures of stimulus control noted in Chapter 3. For example, responding to the mutually entailed relations, and one- and two-node relations that were *different to training*, improved considerably over Chapter 3. For example, high levels of accuracy were observed on the mutually entailed, and one- and two-node relations for participants that passed the experimental task in Experiments 6A and 6B. Similarly, high levels of accuracy were observed for participants in Experiment 8A on these same test relations. A primary difference between Chapters 3 and 4 centred on the protocols employed to examine derived comparative responding. In Chapter 3, a conditional discrimination was employed, while in Chapter 4, a variant of the RCP and constructed-response protocol was employed. Thus, the introduction of the RCP may have helped to overcome some of the potential limitations associated with the conditional discrimination employed in Chapter 3. In addition, findings from Chapter 4 suggested that sufficient exposure to non-arbitrary training trials is necessary to facilitate responding in accordance with derived comparative relations. For example, results from Experiments 7A and 7B revealed that when participants were exposed to a reduced number of non-arbitrary training trials, response accuracy on many test relations (e.g., mutual and combinatorial entailment), returned to the earlier, low levels observed across Experiments 4-5B in Chapter 3. These issues will now be explored in further detail in the following sections.

Linearity

The linearity, or sequential presentation, of arbitrary relational training pairs was varied across the two experiments. Previous research conducted in domains outside behaviour analysis suggests that the order in which arbitrary relational training pairs are presented may influence arbitrary test performance (e.g., Hunter, 1957; Russell et al., 1996). Cognitive models of such performance differences suggest that presenting the training pairs in a linear order allows participants to covertly organise and arrange the stimuli into a unified linear representation or mental model, which is then inspected during testing to make inferences about the novel, nonadjacent stimulus pairs (e.g., Acuna et al., 2002; Sedek & von Hecker, 2004; Williams, Avery, Wooland,

& Heckers, 2012). Behaviour analysts, however, reject such mediational explanations and have instead sought to develop parsimonious accounts of the effects of linearity on performance during tests for derived comparative relations (Gorham et al., 2009; Munnelly et al., 2010; Vitale et al., 2008).

With respect to the current findings, the linearity of training pairs did not have a significant effect on mean overall yield in Experiments 6A (linear) and 6B (non-linear). That is, regardless of the linearity of training pairs, an almost equal number of participants in both experiments passed the experimental task. When performances on linear and non-linear test relations were examined for the participants that passed both experiments, high accuracy was observed across all trial types. Thus, the linearity of the test trials did not affect relational performances at test for participants in Experiments 6A and 6B, which is contrary to previous findings (e.g., Russell et al., 1996). In Experiment 7A, only two participants passed, while seven participants passed Experiment 7B, and thus, a significant difference was observed in terms of mean overall yield. Similar to Experiments 6A and 6B, high accuracy was again observed on both linear and non-linear test trials, for participants that passed. Experiments 8A and 8B also examined the effects of linearity in terms of mean overall yield, but no significant differences were observed between experiments on this measure. For the participants that passed both experiments, again no differences were observed on performance accuracy to either the linear or non-linear test relations.

The current findings therefore extend efforts examining the effects of linearity on relational abilities at test, from a 3- (e.g., Vitale et al., 2008) to 5-member network with adult participants. The current findings however, do not support proposals by other researchers, that the presentation of training pairs in a linear order allows participants to solve test problems with greater ease (e.g., Hunter, 1957; Russell et al., 1996). In the current study, participants displayed high levels of accuracy on both linear and non-linear test relations, regardless of whether they were presented with training pairs in a linear or non-linear order. However, the results from Experiments 7A and 7B did reveal that a large number of participants were unsuccessful in completing both experiments. This finding though may be due to the fact that for a number of participants, weaknesses were observed in their ability to solve mutually entailed problems, and thus, it may not be the features of linearity per se, that contributed to these failures. Therefore, the current findings appear to suggest that although linearity plays an important role in the ability to solve relational problems, there may be other variables that exert stronger control over such performances.

Multiple-exemplar training

An important aim of Experiments 8A and 8B was to examine the potential facilitative effects of a multiple-exemplar training (MET) intervention on the development of arbitrary comparative responding. Such interventions are drawn from principles outlined in the literature on *derived relational responding,* which state that an appropriate history of differential reinforcement across multiple stimulus sets is key to the development of arbitrary relational responding (Hayes et al., 2001b). Findings demonstrated that all participants in Experiment 8A passed the experimental task, while two out of four participants in Experiment 8B.

Despite the fact that all participants in Experiment 8A passed the task, it is difficult to assess the impact of the MET intervention on derived relational performances, as all participants were exposed to the intervention from the beginning of the experiment. On the other hand, participants in a study by Vitale et al. (2008) were only exposed to the MET intervention when tests revealed that participants were deficient on some of the targeted relations. Thus, a potential limitation of the current study is that participants were not assessed on their ability to respond to arbitrary comparative relations before exposure to the intervention. Future studies should seek to expose participants to training interventions, only if relational abilities are found to be deficient. This in turn may help to identify more clearly the conditions that are necessary to facilitate the emergence of derived relational responding. Furthermore, the intervention employed by Vitale et al. differed from that employed in the current study. For example, the intervention employed in the Vitale et al. study, involved converting the deficient arbitrary relations into a non-arbitrary form. That is, the size of the coins that were employed as arbitrary stimuli were altered such that they were no longer identical in size, and thus, the relation between them was no longer arbitrary. In contrast, in the current study, the MET intervention consisted of exposing participants to an additional non-arbitrary training phase at the start of the experiment. Future studies should therefore seek to examine the utility of incorporating interventions that highlight the deficient arbitrary relations in more simpler, non-arbitrary forms.

Relational Completion Procedure

The findings from the current series of experiments also highlight the effectiveness of the RCP and constructed-response format as novel procedures for examining derived comparative relations. With this method, stimuli were presented sequentially from left-to-right on the computer screen and participants were required to "construct" their "relational sentences" in the upper portion of the screen. The comparably high, accurate test performances observed in the Experiments 6A, 6B and 8A, may be partially explained by the response requirements of the RCP. In comparison with the majority of MTS procedures, the RCP allows participants to first complete the relation, and then either evaluate it (by confirming their selection) or initiate a new selection (Dymond & Whelan, 2010). In a traditional MTS task, participants select a comparison by clicking on it in the presence of a particular contextual cue (Dymond & Barnes, 1995; Munnelly et al., 2010; Reilly et al., 2005). Dymond and Whelan (2010) propose that, discriminative control may be enhanced with the RCP, as placing the comparison stimulus on the same level as the sample and the contextual cue moves it away from the rejected stimulus. The authors further propose that, in contrast, discriminative control may be diminished with the top-down method of stimulus presentation seen in the MTS. The same may apply to the findings from the current thesis. For example, the method in which participants made responses to the conditional discrimination paradigm in Chapter 3 was similar to that observed with the MTS protocol in the Dymond and Whelan (2010) study. Thus, if discriminative control is diminished with the top-down method of stimulus presentation, this may potentially account for the failures in stimulus control noted during the arbitrary relational phases in Chapter 3. However, in order to more fully determine whether greater stimulus control may be achieved with the RCP, it may be necessary for future studies examining derived comparative responding, to undertake a direct comparison of the RCP, and the conditional discrimination paradigm.

In addition, Dymond and Whelan (2010) found a facilitative effect for the confirmatory response requirement, with a greater number of participants completing the experiment when they were provided with the opportunity to evaluate and confirm their responses in comparison to those that were not. With respect to the current study, all participants were exposed to the confirmatory response requirement. Across Experiments 6A and 6B, 21 out of 24 (87.5 %) participants passed the experimental task. In Experiments 7A and 7B, a total of 9 out of 24 (37.5 %) participants passed, while across Experiments 8A and 8B, 6 out of 8 (75%) participants passed the experimental task. However, with the current series of experiments, it is difficult to assess the potential facilitative effects associated with the confirmatory response requirement, as it was employed across all experiments. In contrast, Dymond and Whelan (2010) explored the utility of this response option by only exposing half of the participants to experimental conditions involving the confirmatory response requirement. The authors found that, for the seven participants in Conditions 1 to 4 that passed, five received the confirmatory response requirement. Similarly, for the eight participants in Conditions 5 to 8 that passed, seven received the confirmatory response requirement. Thus, the authors concluded that the confirmatory response requirement had a facilitative effect on the emergence of relational responding. However, the current study did not undertake such an investigation, and therefore, future research should seek to manipulate the presence and absence of the confirmatory response requirement.

Feedback thermometer

A noteworthy feature of Chapter 4 was the inclusion of a 'task feedback thermometer' during training and testing phases. The thermometer incremented following correct responses and, because the training criterion involved responding consecutively across a block of trials, it reset to zero when an error was made. During testing phases, the thermometer incremented following each response (correct and incorrect) and was not reset if participants made an error. This additional onscreen feedback was employed during all phases of the experiments as a motivating operation (Michael, 1993) to increase engagement with the task and to make phase progression and task termination reinforcing. Further research should seek to determine the relative effectiveness of the task thermometer feedback during training and testing of multiple stimulus relations.

Touch screen-based responding

The current series of experiments also explored the utility of employing a touch screen-based response system as an alternative to the standard keyboard method of responding. For example, the use of touch screen monitors may maximise responding, as participants are presented with fewer response options, which in turn, may help to minimise errors. Furthermore, as participants are able to respond to trials with greater ease, this may also be beneficial when response times are of importance. It has also been reported that increased interactivity produces increased learning (e.g., Fletcher, 1990; Kritch & Bostow, 1998; Schaffer & Hannafin, 1986). Therefore, incorporating touch screen technology during learning tasks may bear relevance to the applied setting, such as the classroom. For example, Dube et al. (1991) outline a number of advantages associated with the implementation of a computer-based programmed instruction for teaching spelling in the classroom. One such advantage is that the computer produces feedback on a continuous basis during both correct and incorrect learning trials, which in turn, optimises the teacher or assistant's time (e.g., Connors, Caruso, & Detterman, 1986). In addition, the program can be modified to suit the individual. For instance, if a participant displays weak responding on some learning or test trials, then the program can be tailored to re-cycle participants through training and test phases, without the need for teacher assistance. This method of task presentation may also help to circumvent problems associated with fatigue and inattention, which are often encountered during learning tasks. Thus, the current RCP experimental paradigm may be well suited to the classroom setting.

Experimenter-delivered instructions and the RCP

It may be possible to adapt the current protocol to examine the facilitative effects of a relational training and testing intervention that relied solely on experimenter-delivered verbal instructions. For instance, if participants were exposed to an extensive history of exemplar training across multiple stimulus sets, in which they were explicitly informed of the relative value of each stimulus, it may be possible to examine the emergence of derived comparative responding with the presentation of untrained, novel stimulus combinations at test. Indeed, previous studies have found such interventions to be effective in generating symmetrical and comparative responding in both typically developing children, and children with autism, when these repertoires were found to be deficient (e.g., Barnes-Holmes, Barnes-Holmes, Roche, & Smeets, 2001a, 2001b; Berens & Hayes, 2007; Gorham et al., 2009). For instance, the procedures used to train and test derived comparative relations in the Gorham et al. (2009) study may be comparable to the method of stimulus presentation observed in the current study. At the start of each trial in the Gorham et al. study, the researcher placed two coins side by side on the table and told the child that they were going to play a game. The researcher then told the child to imagine that it was their birthday and that they were going to the shop to buy some sweets. Next the researcher instructed the child that "This coin (researcher pointed to the coin on the left) buys more sweets than this coin (researcher pointed to the coin on the right)", followed by the question, "Which would you take to buy as many sweets as possible?" A correct response in this instance involved the child selecting the coin on the left to buy as much sweets as possible (Gorham et al., 2009). Thus, the fact that Gorham et al. established derived comparative responding in this manner for both typically developing children, and children with autism, highlights the utility of experimenter-delivered verbal instructions and the sequence in which stimuli are presented, in generating this type of behaviour. Future research should therefore seek to undertake such an examination with the RCP, with young children, and individuals that lack sophisticated verbal repertoires.

In conclusion, the current series of experiments investigated the effectiveness of an alternative training and testing paradigm in examining "More-than" and "Less-than" relations in adult participants. For the most part, the RCP along with the constructed-response protocol, were successful in establishing responding to derived comparative relations. In addition, two experiments examined the utility of a non-arbitrary intervention in facilitating responding to arbitrary comparative relations (Experiments 8A and 8B). However, the impact of the intervention on relational performances remains unclear and thus, future studies should seek to clarify this issue. Despite this, the RCP has the potential to examine a range of multiple stimulus relations (e.g., "Before" and "After" and "Spatial"), and future research should seek to examine its utility with other population samples including young children and individuals that lack sophisticated verbal repertoires.

Chapter 5

The Transformation of Discriminative Functions in Accordance with a 5-member Comparative Relational Network

Dymond and Barnes (1995) report evidence for the transformation of self-discriminative functions in accordance with the relational frames of "Same", "More-than" and "Less-than", while Whelan et al. (2006) demonstrated the transformation of "More-than" and "Less-than" consequential functions to a 7-member relational network. More recently, Dougher et al. (2007) examined the transformation of discriminative functions with a 3-member relational network in which the stimuli were ranked in terms of size. In this study, participants were first exposed to a relational training phase in which they were presented with one of three arbitrary sample stimuli (A, B and C) in the top portion of the computer screen, and three comparison stimuli in the bottom portion of the screen (Experiment 1). The comparison stimuli were physically similar but differed in terms of size (e.g., small, medium and large). The purpose of this phase was to train participants to select the smallest comparison in the presence of sample A, the medium comparison in the presence of sample B, and the largest comparison in the presence of sample C. Once participants met criterion at testing, they were exposed to a bar press training and test phase. During this phase of the experiment, participants were initially trained to press the spacebar at a steady rate to the medium (B) stimulus. Once participants pressed the spacebar at a constant rate for three consecutive trials in the presence of stimulus B, they were exposed to a test phase in which the small (A) and large (C) comparison stimuli were now presented alongside the middle (B) stimulus. Results demonstrated that participants pressed slower to stimulus A and faster to stimulus C, than they did to the middle stimulus B. During a subsequent phase of the experiment, participants were exposed to respondent conditioning with stimulus B, and testing with stimuli A and C. Thus, stimulus B was paired with a mild shock, and changes in skin conductance were employed as the dependent variable. Dougher et al. (2007) found that 6 out of 8 participants demonstrated smaller skin conductance changes to stimulus A, and larger changes to stimulus C, than to stimulus B.

The phenomena just described, the transformation of stimulus functions, outlines how the function attached to one stimulus in a relational network alters the functions of other members in the same network, in accordance with the derived relation between the stimuli (e.g., Dougher & Markham, 1994, 1996; Dymond & Rehfeldt, 2000; Hayes, 1991). For example, a function trained to the middle-ranking stimulus C in a 5-member relational network (E>D>C>B>A), may result in participants responding "more" to the stimuli ranked higher in the network (i.e., D and E), and "less" to the stimuli ranked lower in the network (i.e., A and B), in the absence of any further training. Previous findings have demonstrated the transformation of functions in accordance with the relational frames of "More-than" and "Less-than" (e.g., Dougher et al., 2007; Dymond & Barnes, 1995; Whelan et al., 2006), "Same" and "Opposite" (e.g., Dymond et al., 2007, 2008; Whelan & Barnes-Holmes, 2004), and also equivalence relations (e.g., Dougher, Auguston, Markham, Greenway, & Wulfert, 1994; Dougher, Perkins, Greenway, Koons, & Chiasson, 2002; Roche & Barnes, 1997). Functions shown to transform include self-discrimination (e.g., Dymond & Barnes, 1994, 1995), consequential (e.g., Whelan et al., 2006), discrimination (e.g., Dougher et al., 2007), avoidance (e.g., Auguston & Dougher, 1997; Dymond et al., 2007, 2008), and respondent elicitation and extinction (e.g., Dougher et al., 1994; Dougher & Markham, 1996; Roche & Barnes, 1997; Roche, Barnes-Holmes, Smeets, Barnes-Holmes, & McGeady, 2000).

With respect to the Dougher et al. (2007) study, the authors argue that participants in their study were successful in displaying the transformation of discriminative functions. However, some inconsistencies in their training procedures have been noted. For example, Stewart and McElwee (2009) note that the authors refer to their training procedures as arbitrary matching-to-sample, whereas others (e.g., Berens & Hayes, 2007) refer to similar procedures as non-arbitrary relational training. With respect to the Dougher et al. study, in Experiment 2, stimulus A was employed as a sample stimulus to establish an arbitrary size ranking between four circles that were the same size, but differed in terms of colour. The authors then employed one of the "middle" circles to train participants to respond at a steady rate of bar pressing, before introducing the other circles. Dougher et al. (2007) found that 5 out of 6 participants pressed the spacebar slower to the "smaller" circle, and faster to the "larger" circle, than to the "middle" circle. Thus, similar to their findings in Experiment 1, the authors argue that the relational training procedures were successful in transforming the functions of one member of the relational network (A), so that it could be used to establish a rank ordering between the other coloured circles. That is, the relational training procedures transformed the functions of stimuli A, B and C in Experiment 2, so that participants ranked A as the "smallest", B as the "middle" and C as the "largest". However, as the authors note, it may have been possible that the transformation of discriminative functions observed in both Experiments 1 and 2 may not have been due to the derived relations among the sample stimuli (p.191). That is, the functions that stimuli A and C acquired in Experiment 1 may not have resulted in their derived relation to stimulus B, but as a result of their non-arbitrary association with the smallest and largest comparisons (Dougher et al., 2007).

Furthermore, Stewart and McElwee (2009) state that the authors define *arbitrary* on the basis that there is no consistent physical relation between the sample (contextual cue), and the comparison stimuli (e.g., non-arbitrary stimuli). However, in order to meet the definition of *arbitrary* in accordance with RFT, Stewart and McElwee state that the stimulus presented at the top of the computer screen functions as a contextual cue, and not as a sample stimulus involved in a relation with other comparison stimuli, to control relational responding. Thus, Stewart and McElwee (2009) conclude that Dougher et al. (2007) were successful in generating non-arbitrary, but not arbitrary, comparative relational responding between the comparisons under the control of contextual cues. A subsequent experiment (Experiment 3) by the authors however, did demonstrate that the relational training procedures could establish derived relations between arbitrary stimuli in a 3-member relational network. However, Dougher et al. did not undertake an investigation of the transformation of discriminative functions in accordance with this network, and thus, it remains to be seen whether the predicted patterns of performance would emerge with those stimuli.

Current Experiments

The current study aimed to replicate and extend Dougher et al.'s (2007) findings from a 3- to 5-member relational network. In addition, the current study also aimed to address a limitation noted with the relational training and testing procedures employed in the Dougher et al. study, and demonstrate the transformation of discriminative function in accordance with the *arbitrary* relation between stimuli. In Experiment 9A, participants were first exposed to non-arbitrary relational training and testing to establish the contextual functions of MORE-THAN and LESS-THAN for two abstract images. Arbitrary relational training and testing followed, where participants received training on "More-than" baseline relations (e.g., B>A, C>B, D>C and E>D), followed by testing with a combination of "More-than" and "Less-than" relations (e.g., C>A and A<C). Next, participants were exposed to a bar-press training phase, which aimed to train a function to the middle stimulus (C) in the relational network. A test for the transformation of discriminative functions followed, in which participants were exposed to probe trials involving all members of the relational network (A-B-C-D-E). In addition, this test phase differed from Dougher et al.'s (2007) in that the test stimuli were not presented in a fixed order for one time each, but instead, the members of the relational network were presented in a quasi-random order, three times each within a test block. It was predicted that participants would respond "more" to the stimuli higher in the network (e.g., D and E), and "less" to the stimuli lower in the network (e.g., A and B).

Experiments 9B-9D also sought to explore the transformation of discriminative functions to a 5-member arbitrary relational network.

Experiment 9A

Method

Participants

Nine participants, five male and four female, ranging in age from 20 to 36 years (M_{age} = 26.00, *SD* = 6.24), were recruited through personal contacts and notice-board announcements at Swansea University. Participants were paid £8 on completion of the study.

Apparatus and Setting

The experiment was conducted in an experimental room (2 X 3 metres) in the Psychology Department at Swansea University. All training and test trials were presented on a 16-inch display screen by a programme written in Visual Basics.NET.

Materials and Stimuli

The same arbitrary images as those employed in Experiments 6A-8B were employed as the contextual cues for MORE-THAN and LESS-THAN in the current study (see Figure 4.1). In addition, the non-arbitrary stimulus sets employed in the current study were the same as those used in Experiments 6A-8B (see Appendix 5 for a full list of the non-arbitrary stimulus sets).

For the arbitrary relational training and testing phases, five abstract images (see Figure 5.1) were used to generate a 5-member linear relational network (A-B-C-D-E).

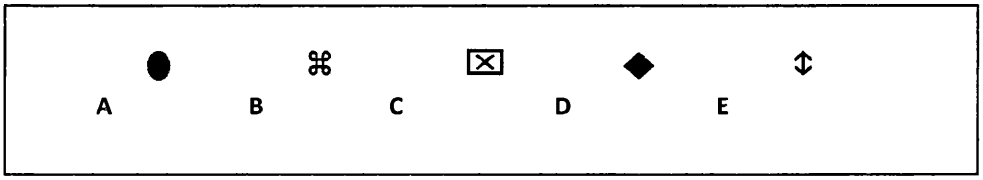


Figure 5.1. The five abstract images employed during arbitrary relational training and testing, and are labelled A, B, C, D and E in the interests of clarity (Note: participants were never exposed to these labels).

Procedure

The general procedural sequence was based on those employed by Dymond and Whelan (2010), and was as follows: *Phase 1A: Constructed-Response Non-arbitrary Relational Training* and *Phase 1B: Constructed-Response Non-arbitrary Relational Testing; Phase 2A: Constructed-Response Non-arbitrary Relational Training (multiple-exemplar training)* and *Phase 2B: Constructed-Response Non-arbitrary Relational Testing; Phase 3: Constructed-Response Arbitrary Relational Training; Phase 4: Arbitrary Relational Test 1; Phase 5: Arbitrary Relational Test 2* and *Phase 6:* The procedure employed in this phase of the experiment was based on Dougher et al. (2007), and was as follows: *Bar press training with stimulus C and testing with stimuli A, B, C, D and E..*

For Phases 1A-5 in the current study, the training and testing procedures were the same as those employed in Experiment 8A (see Figure 4.4 in Chapter 4 for an overview of the arbitrary relational training and test trials that participants in Experiment 9A were exposed to). In addition, mastery criterion during all training and testing phases was identical to Experiment 8A.

Phase 6: Bar press training with stimulus C and testing with stimuli A, B, C, D and E. In this phase of the experiment, participants were trained to press the spacebar on the computer keyboard at a steady rate when stimulus C was presented onscreen. This phase began with the following instructions onscreen:

> During this part of the experiment, a symbol from the previous part will appear in the centre of the computer screen. When you see the symbol, your task is to repeatedly press the spacebar on the keyboard for the entire time the symbol is presented. Do not just hold down the spacebar; press it repeatedly. Your task is to try and obtain a steady rate of spacebar presses in the presence of this symbol. Each time you press the spacebar, a mark will appear on the bottom of the computer screen. There is no feedback other than this during this phase of the experiment. The same symbol will appear repeatedly until you press the spacebar at a steady rate in the presence of this symbol. Later, a number of other symbols from the previous parts will then be presented. Again, your task is to press the spacebar, at a rate you feel appropriate, for each new symbol. Please ask the experimenter if you have any questions whatsoever.

Clicking on the OK button removed the instructions and signalled the start of Phase 6. This training phase aimed to establish a steady rate of bar pressing in the presence of stimulus C, which was presented in the centre of the computer screen. Each time a participant pressed the spacebar, a dash appeared on the bottom of the screen to signal the number of bar presses made. Participants were first exposed to a practice trial in which the experimenter demonstrated how many bar presses they were required to make to stimulus C during that trial. For example, the experimenter pressed the spacebar at a steady rate of one bar press per second, for the duration of 30 s. The program was then restarted and the experimenter instructed the participant to respond to the stimulus in exactly the same manner. In order to meet criterion on this test phase, participants were required to make 30 bar presses (+/- 10%) to stimulus C, for three consecutive trials. Training trials were repeated until this criterion was achieved. Once met, participants were immediately exposed to a test phase, in which stimuli from the previous phase of the experiment were presented alongside stimulus C. The five test stimuli were presented in a quasi-random order, for a total of three times each, resulting in a total of fifteen test trials. Again, each stimulus was presented for the duration of 30 s and no criterion was in place during testing. Once all fifteen test trials were presented, the experiment ended and participants were asked to report to the experimenter.

Results and Discussion

Of the nine participants that took part in Experiment 9A, all nine successfully completed both arbitrary relational test phases, and required between 1 and 2 (M = 1.22, SD = .44) exposures to testing to do so (see Table 5.1). In addition, all nine participants were exposed to training and testing for the transformation of discriminative functions. Table 5.1 displays the trials to criterion for participants during both constructed-response non-arbitrary relational training phases and constructed-response arbitrary relational training. All nine participants required between 10 and 24 (CR-NARB1: M = 14.00, SD = 5.31; CR-NARB2: M = 10.00, SD = .00) trials to reach criterion during the constructed-response non-arbitrary relational training phases. The number of trials required to achieve constructed-response arbitrary relational training criterion ranged between 24 and 36 (M = 32.73, SD = 5.61).

As the current experiment was concerned with the transformation of discriminative functions in accordance with the relational frames of "More-than" and "Less-than" to a 5-member relational network, the results section will not discuss performances for participants during both arbitrary relational test phases. Instead, a summary of performances can be seen in Table 5.1. The remainder of the results section will consider performances for participants during Phase 6 (Bar press training with stimulus C and testing with stimuli A, B, C, D and E) of the experiment.

Table 5.1
Trials to criterion for the All-More training group during constructed-response non-arbitrary and arbitrary relational training in Experiment 9A. Also displayed is individual data for the All-More group.

	Phases 1A&2A: CR-NARB Relational Training		Phase 3: CR-Arbitrary Relational Training	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant				B	ME	B	CE1	CE2
1	11	10	36	16/16	16/16	16/16	23/24	16/16
2	11	10	24	16/16	16/16	16/16	24/24	16/16
3	10	10	36	16/16	16/16	16/16	24/24	16/16
4	15	10	36	15/16	15/16	12/16	14/24	11/16
	10	10	36	16/16	16/16	15/16	23/24	16/16
5	11	10	36	16/16	16/16	16/16	24/24	16/16
6	23	10	36	11/16	11/16			
	10	10	24	16/16	16/16	16/16	24/24	16/16
7	18	10						
		10	36	16/16	16/16	14/16	24/24	16/16
8	11	10	36	16/16	16/16	16/16	24/24	16/16
9	24	10	24	15/16	16/16	16/16	24/24	16/16

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2.

Bar press training with stimulus C, and testing with stimuli A, B, C, D and E

In order to meet training criterion during Phase 6, participants were first required to make 30 spacebar presses (+/- 10%) across three consecutive exposures to stimulus C. Results demonstrated that participants required between 3 and 53 (M = 22.67, SD = 16.07) training trials to meet criterion. Participants were then exposed to testing with stimuli A, B, C, D and E, which were presented three times each, in a quasi-random order. Results for this part of the experiment are discussed with respect to the number of bar presses participants made to each stimulus. That is, as participants were exposed to each of the five stimuli, three times during testing, data is displayed and discussed for the number of bar presses (per 30 s) participants made to each stimulus on the first, second and third time they encountered these stimuli.

Participants who demonstrated the transformation of discriminative functions

A total of four (P1, P3, P6 and P7) out of nine participants demonstrated the transformation of "More-than" and "Less-than" discriminative functions to the 5-member relational network, following arbitrary relational training and testing (see Figure 5.2).

P1. On his/her first exposure to each test stimulus, P1 pressed the spacebar "less" to stimuli A, B and D, and "more" to stimulus E than to stimulus C. However, on his/her second and third exposures, P1 pressed "less" to stimuli A and B and "more" to stimuli D and E than to stimulus C, and thus, successfully demonstrated the transformation of functions to all five test stimuli.

P3. Across all three exposures to the test stimuli, P3 pressed the spacebar "less" to stimulus A, and "more" to stimuli B, D and E than to stimulus C. Thus, P3 demonstrated transformation of functions to stimuli A, D and E, but not to stimulus B.

P6. On his/her first exposure to the test stimuli, P6 pressed the spacebar "more" to stimuli A, B and E than to stimuli C and D. However, on his/her second and third exposures, P6 pressed "less" to stimuli A and B and "more" to stimuli D and E than to stimulus C, thus demonstrating the transformation of functions to all test stimuli.

P7. Across all three exposures to the test stimuli, P7 demonstrated consistent transformation of functions to all test stimuli.

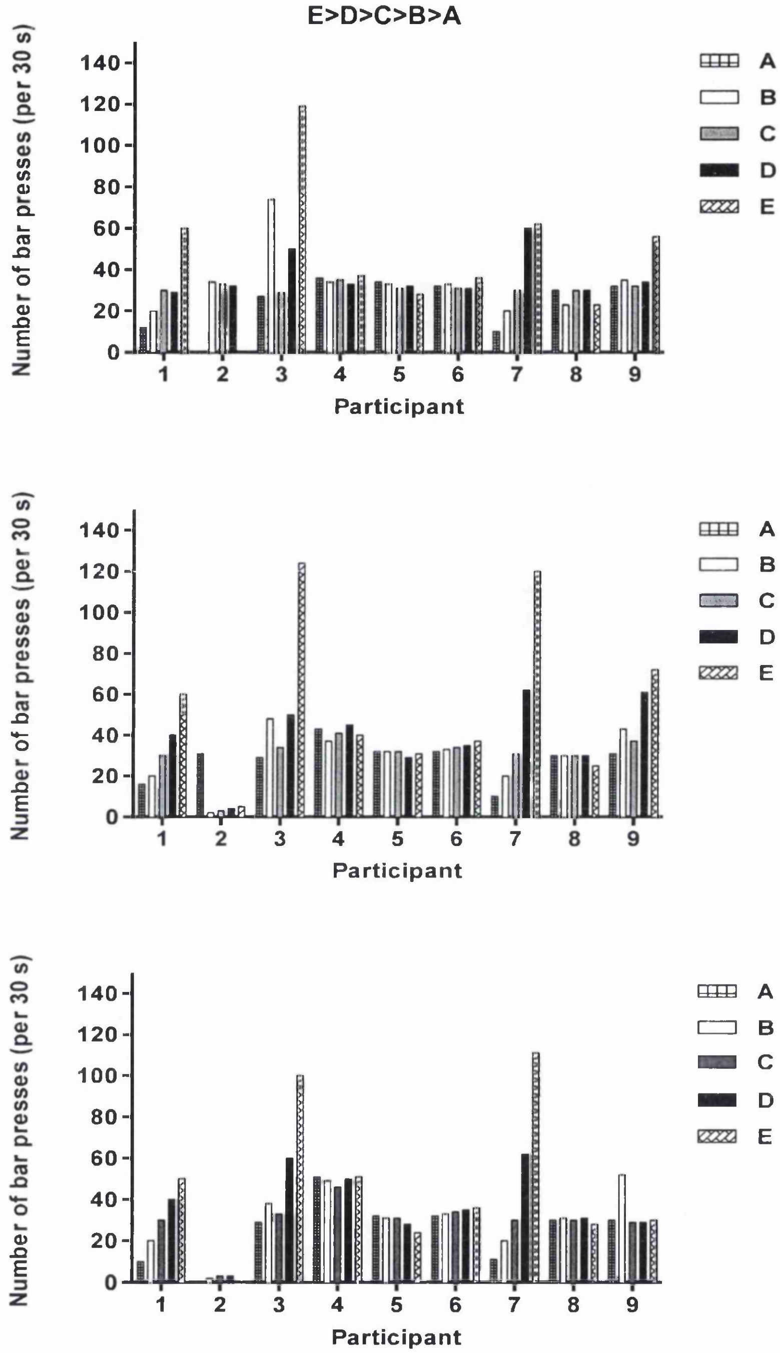


Figure 5.2. Experiment 9A. The number of bar presses made per 30 s to stimuli A, B, C, D and E, for the nine experimental participants in the bar press testing phase of Experiment 9A. The upper figure refers to the number of bar presses (per 30 s) that participants made the first time they encountered each stimulus during testing, the middle figure refers to the number of bar presses made the second time they encountered the test stimuli, while the bottom figure refers to number of bar presses made the third and final time they encountered these stimuli.

Participants who failed to demonstrate the transformation of discriminative functions

P2. On his/her first exposure to the test stimuli, P2 pressed the spacebar "more" to stimulus B and "less" to stimulus D than he/she did to stimulus C. P2 did not make any bar presses to stimuli A and E. On his/her second exposure, P2 pressed "more" to stimuli A, D and E, and "less" to stimulus B than to stimulus C. On his/her third and final exposure to the test stimuli, P2 pressed "less" to stimulus B and "more" to stimulus D than to stimulus C. Similar to their first test exposure, P2 did not make any bar presses to stimuli A and E.

P4. On his/her first exposure to the test stimuli, P4 pressed the spacebar "more" to stimuli A and E and "less" to stimuli B and D than to stimulus C. On his/her second exposure, P4 pressed "more" to stimuli A, D and E and "less" to stimulus B than to stimulus C. On his/her final exposure to the test stimuli, P4 pressed "more" to stimuli A, B, D and E than to stimulus C.

P5. On his/her first and third exposures to the test stimuli, P5 pressed the spacebar "more" to stimuli A, B and D, and "less" to stimulus E than to stimulus C. On his/her second exposure, P5 pressed "more" to stimuli A and B and "less" to stimuli D and E than to stimulus C.

P8. On his/her first exposure to the test stimuli, P8 pressed the spacebar the same number of times to stimuli A, C and D and "less" times to stimuli B and D. On his/her second exposure, P8 made the same number of bar presses to stimuli A, B, C and D and "less" to stimulus E. Finally, on his/her third exposure to the test stimuli, P3 made the same number of bar presses to stimuli A and C and slightly "more" to B and D than to stimulus C.

P9. On his/her first and second exposures to the test stimuli, P9 pressed the spacebar "more" to stimuli B, D and E, and "less" to stimulus A than to stimulus C. On his/her third and final exposure to the test stimuli, P9 pressed "more" to stimuli A, B and E and slightly "less" to stimuli C and D.

The results of Experiment 9A demonstrated that there was considerable variation across participants during tests for the transformation of functions. For instance, only four out of nine participants demonstrated the predicted patterns of performance. That is, these participants pressed the spacebar "less" to stimuli A and B, and “more” to stimuli D and E than to stimulus C. On closer inspection, P7 was the only participant, across all three exposures to the test stimuli, to consistently press “less” to stimuli A and B, and “more” to stimuli D and E than to stimulus C. On the other hand, P1 and P6 demonstrated consistent transformation on their second and third exposures, but not their first. P3 also demonstrated transformation to four of the test stimuli (A, C, D and E) across three exposures. However, P3 did not demonstrate the predicted transformation across the same three exposures to stimulus B.

In addition, a number of participants failed to display the predicted patterns of performance. These findings are in contrast to Dougher et al.’s (2007) study, in which all eight experimental participants in Experiment 1, and five out of six participants in Experiment 2 responded in accordance with the rank ordering of the 3-member network (A<B<C) during transformation tests. Although the reasons for the current findings remain unclear, one potential cause may have been that the relational network was not well established for these participants. However, upon closer inspection, four of the five participants (P2, P5, P8 and P9) passed both arbitrary relational test phases on their first exposure to testing, achieving perfect accuracy on all test relations. P5 on the other hand, required two exposures to training and testing, but displayed near perfect accuracy on all test relations on their second exposure to testing. Thus, the derived relational network appears to have been clearly established for the participants that failed to demonstrate the transformation of “More-than” and “Less-than” discriminative functions, and therefore, further research is needed to determine the factors affecting the emergence of this type of responding.

With respect to the current study, the method in which the stimuli were presented during transformation tests differed from Dougher et al. (2007). For example, in the current study, the stimuli were presented in a quasi-random order, three times each within a test block. In contrast, in the Dougher et al. study, the test stimuli were presented once each in a fixed stimulus sequence (e.g., A-B-C). Thus, the sequence of stimulus presentation during transformation tests may be an important issue as a larger number of participants in the Dougher et al. (2007) study displayed the predicted patterns of performance.

Experiment 9B

Experiment 9B examined whether predicted patterns of transformation would emerge (i.e., press "more" to D and E and "less" to A and B, than to C) following training with only "Less-than" baseline relations. Therefore, participants in Experiment 10B were exposed to arbitrary relational training with "Less-than" relations (A<B<C<D<E), followed by arbitrary relational testing with both "More-than" and "Less-than" relations. The bar-press training and testing phase remained the same as in Experiment 9A.

Method

Participants

Eight participants, three male and five female, ranging in age from 19 to 30 years (M_{age} = 21.50, *SD* = 3.70), were recruited through personal contacts and notice-board announcements at Swansea University. Participants were paid £8 on completion of the study.

Procedure

The procedure for Experiment 9B was identical to Experiment 9A, with the exception of arbitrary relational training. During this phase of the experiment, participants were now presented with All-Less training pairs (A<B, B<C, C<D and D<E), as opposed to the All-More training pairs (B>A, C>B, D>C and E>D) presented in Experiment 9A. Mastery criterion during all training and testing phases was the same as Experiment 9A.

Results and Discussion

Of the eight participants that took part in Experiment 9B, all eight successfully completed both arbitrary relational test phases, and required between 1 and 2 (*M* = 1.25, *SD* = .47) exposures to testing to do so (see Table 5.2). In addition, all eight participants were exposed to training and testing for the transformation of discriminative functions.

Table 5.2 displays the trials to criterion for participants during both constructed-response non-arbitrary relational training phases and constructed-response arbitrary relational training. All eight participants only required between 10 and 12 (CR-NARB1: *M* = 10.82, *SD* = .60; CR-NARB2: *M* = 10.00, *SD* = .00) trials to reach criterion during the constructed-response non-arbitrary relational training phases. The number of trials required, to achieve constructed-response arbitrary relational training criterion, ranged between 24 and 60 (M = 27.60, SD = 12.71). Again, the current experiment was concerned with the transformation of discriminative functions in accordance with the relational frames of "More-than" and "Less-than" to a 5-member relational network, and thus, the results section will not discuss performances for participants during both arbitrary relational test phases. Instead, a summary of performances can be seen in Table 5.2.

Table 5.2
Trials to criterion for the All-Less training group during constructed-response non-arbitrary and arbitrary relational training in Experiment 9B. Also displayed is individual data for the All-Less group.

	Phases 1A&2A: CR-NARB Relational		Phase 3: CR-Arbitrary Relational	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training		Training	B	ME	B	CE1	CE2
1	11	10	24	16/16	16/16	16/16	24/24	16/16
2	12							
	10	10	60	16/16	16/16	13/16	19/24	13/16
	10	10	24	16/16	16/16	16/16	24/24	16/16
3	11	10	24	16/16	16/16	16/16	24/24	16/16
4	11	10	36	16/16	16/16	16/16	24/24	16/16
5	11	10	24	16/16	14/16			
	10	10	12	16/16	16/16	11/16	24/24	16/16
6	11	10	24	16/16	16/16	16/16	24/24	16/16
7	11	10	24	16/16	16/16	16/16	24/24	16/16
8	11	10	24	16/16	16/16	15/16	24/24	15/16

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2.

Bar press training with stimulus C, and testing with stimuli A, B, C, D and E

Similar to Experiment 9A, participants were first required to make 30 spacebar presses (+/- 10%), across three consecutive exposures to stimulus C. Results demonstrated that participants required between 3 and 31 (M = 7.63, SD = 9.61) trials to reach criterion. Participants were then exposed to testing in the same manner as Experiment 9A.

Participants who demonstrated the transformation of discriminative functions

Half of the participants (P1, P2, P3 and P8) that took part in Experiment 9B demonstrated the transformation of "More-than" and "Less-than" functions to the 5-member relational network.

P's 1, 2, 3 and 8. Across all three exposures to the test stimuli, P's 1, 2, 3 and 8, demonstrated consistent transformation of functions to all test stimuli. That is, all four participants pressed the spacebar "more" to stimuli D and E and "less" to stimuli A and B than to stimulus C (see Figure 5.3).

Participants who failed to demonstrate the transformation of discriminative functions

P4. On his/her first and second exposures to the test stimuli, P4 pressed the spacebar "more" to stimuli A and B and "less" to stimuli D and E than to stimulus C. On his/her third and final exposure to the test stimuli, P4 pressed "more" to stimuli A and B and "less" to stimulus E than to stimuli C and D. Thus, across two exposures to the test stimuli, P3 demonstrated the transformation of stimulus functions to the five members of the relational network, but in the opposite direction to the predicted pattern of performance.

P's 5 and 7. Across all three exposures to the test stimuli, P5 and P7 responded almost equivalently to all test stimuli. For example, P5 pressed the spacebar 30 times to stimuli C, D and E, and between 30 and 31 times to stimuli A and B, across all three exposures to testing. Similarly, P7 pressed the spacebar 30 times to stimuli B and E and between 29 and 30 times to stimuli A, C and D, across all exposures to testing. Thus, P5 and P7 did not demonstrate the predicted transformation of "More-than" and "Less-than" functions to the 5-member relational network.

P6. On his/her first and third exposures to the test stimuli, P6 pressed the spacebar the same number of times to stimuli B and C and less times to stimuli A, D and E. On his/her second exposure to the test stimuli, P6 pressed "more" to stimuli B and E and "less" to stimuli A and D than to stimulus C.

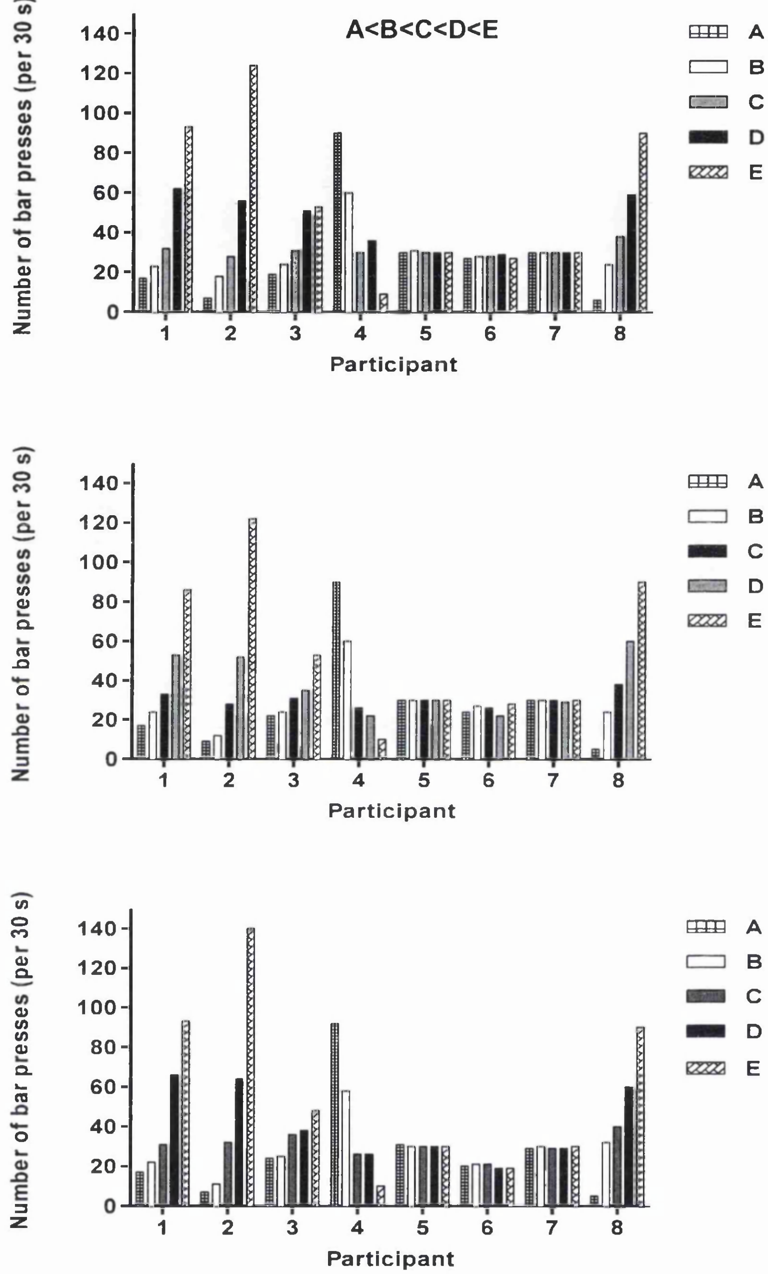


Figure 5.3. Experiment 9B. The number of bar presses made per 30 s to stimuli A, B, C, D and E, for the nine experimental participants in the bar press testing phase of Experiment 9B. The upper figure refers to the number of bar presses (per 30 s) that participants made the first time they encountered each stimulus during testing, the middle figure refers to the number of bar presses made the second time they encountered the test stimuli, while the bottom figure refers to number of bar presses made the third and final time they encountered these stimuli.

In summary, and similar to Experiment 9A, there was considerable variation in participants' abilities to demonstrate the transformation of "More-than" and "Less-than" discriminative functions to the 5-member relational network in Experiment 9B. For example, only half of the participants (P1, P2, P3 and P8) demonstrated the predicted transformation of functions. However, all four participants displayed consistent transformation across all exposures to the test stimuli. On the other hand, participants that failed to demonstrate the predicted patterns of performance, varied considerably in their results. For example, P4 demonstrated the transformation of "More-than" and "Less-than" functions, but in the opposite direction to the established network. In addition, P5 and P7 responded almost equivalently to all test stimuli, while P6 responded the same number of times to stimuli B and C and "less" times to stimuli A, D and E.

The results of both Experiments 9A and 9B taken together, demonstrate that only half of the participants were successful in responding in accordance with the pre-established 5-member relational network during transformation tests. Although the reasons for this remain unclear, and as previously mentioned, one possible reason may lie with the method and sequence in which the stimuli were presented during transformation tests. For example, participants in both studies were exposed to three presentations of each of the five test stimuli, in a quasi-random order, and in contrast, participants in the Dougher et al. (2007) study were exposed to the test stimuli once each, in a fixed order. Therefore, Experiments 9C and 9D sought to explore whether presenting the five members of the relational network in a fixed sequence would facilitate the emergence of this type of responding.

Experiment 9C

Based on findings of Dougher et al. (2007), Experiment 9C sought to examine the potential facilitative effects of presenting the five members of the relational network in a fixed order during transformation tests. Following training with stimulus C, participants in Experiment 9C were presented with the test stimuli in the following sequence: C, C, A, B, C, D and E.

Method

Participants

Nine participants (5 experiment and 4 control), five male and four female, ranging in age from 20 to 33 years (M_{age} = 26.00, SD = 6.24), were recruited through personal contacts at Swansea University.

Procedure

The general procedure for Phases 1A-5 was identical to those employed in Experiments 9A and 9B. In addition to the experimental participants, four control participants were exposed to only training and testing for the transformation of discriminative "More-than" and "Less-than" functions, and thus, these participants were not exposed to non-arbitrary or arbitrary relational training and testing. The main difference between Experiment 9C and Experiments 9A and 9B was during Phase 6, in which participants were now exposed to the test stimuli in a fixed stimulus order, once each, and not in the random order observed in Experiments 9A and 9B.

Phases 1A-5. All phases and mastery criterion were identical to Experiments 9A and 9B.

Phase 6. Bar press training and testing. The bar press training part of this phase of the experiment was identical to Experiments 9A and 9B. However, during testing, the stimuli were now presented in a fixed order. That is, participants were exposed to the stimuli in the following sequence: C, C, A, B, C, D and E. Again, no criterion was employed for this test phase and participants were only exposed to this fixed test order once.

Results and Discussion

Of the five experimental participants that took part in Experiment 9C, four (E1, E2, E4 and E5) successfully completed both arbitrary relational test phases, and required between 1 and 2 (M = 1.50, SD = .58; see Table 5.3) exposures to testing to do so. One participant (E3) failed to meet criterion on the arbitrary relational Test 1 (mutual entailment), and their data is therefore excluded from further analysis. In addition, four control participants (C1, C2, C3 and C4) took part in Experiment 9C, but were not exposed to non-arbitrary or arbitrary relational training and testing. Instead, these participants were exposed only to training and testing for the transformation of "More-than" and "Less-than" discriminative functions.

Table 5.3 displays the trials to criterion for participants during both constructed-response non-arbitrary relational training phases and constructed-response arbitrary relational training. Participants that passed the experimental task only required between 10 and 15 (CR-NARB1: M = 12.00, SD = 2.37; CR-NARB 2: M = 10.00, SD = .00) trials to reach criterion during the constructed-response non-arbitrary relational training phases. The number of trials required, to achieve constructed-response arbitrary relational training criterion, ranged between 12 and 36 (M = 26.00, SD = 9.03) trials.

Table 5.3
Trials to criterion for experimental participants during constructed-response non-arbitrary and arbitrary relational training in Experiment 9C. Also displayed is individual data for the five experimental participants.

	Phases 1A&2A: CR-NARB Relational		Phase 3: CR-Arbitrary Relational	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training		Training	B	ME	B	CE1	CE2
1 (All-More)	11	10	24	15/16	14/16	6/16	1/24	1/16
	10	10	24	16/16	16/16	16/16	21/24	15/16
2 (All-Less)	15	10						
		10	36	14/16	16/16	16/16	24/24	16/16
3 (All-More)*	15	10	60	16/16	1/16			
	10	10	12	16/16	0/16			
	10	10						
	10	10	12	16/16	0/16			
		10	12	16/16	0/16			
4 (All-Less)	15	10	36	8/16	15/16			
	10	10	12	16/16	16/16	15/16	24/24	13/16
5 (All-More)	11	12	24	16/16	16/16	16/16	24/24	16/16

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2. * refers to participants who failed to complete the experiment.

A summary of performance accuracy during Test 1 (mutual entailment) and Test 2 (combinatorial entailment) for participants that passed the experimental task can be seen in Table 5.3. During the bar press training part of phase 6 in Experiment 9C, and similar to Experiments 9A and 9B, participants were first required to make 30 spacebar presses (+/- 10%) across three consecutive exposures to stimulus C. Results demonstrated that participants required between 3 and 20 (M = 10.50, SD = 7.94) trials to reach criterion.

Of the four experimental participants that were exposed to tests for the transformation of "More-than" and "Less-than" discriminative functions, none demonstrated the predicted transformation of functions to the five members of the relational network.

Participants who failed to demonstrate the transformation of discriminative functions

E1. On his/her first and only exposure to the test stimuli, E1 pressed the spacebar "more" to stimuli A and D and "less" to stimuli B and E, than to stimulus C (see Figure 5.4). Thus, E1 demonstrated the predicted transformation of functions to stimuli B and D, but not to stimuli A and E.

E2. On his/her exposure to the test stimuli, E2 pressed the spacebar "more" to stimuli A and B and "less" to stimuli D and E, than to stimulus C (see Figure 5.4). Thus, E2 demonstrated the transformation of "More-than" and "Less-than" functions, but in the opposite direction of the established network.

E4. On his/her exposure to the test stimuli, E4 pressed the spacebar "less" to stimuli A, B and D, than to stimulus C. E4 also pressed the same number of times to stimuli E and C (see Figure 5.4). Thus, E4 demonstrated the transformation of discriminative functions to stimuli A and B, but not to stimuli, D and E.

E5. On his/her exposure to the test stimuli, E5 pressed the spacebar "more" to stimuli A, B, D and E, than to stimulus C (see Figure 5.4). However, as E5 pressed "more" to stimulus D than to stimulus E, he/she failed to display consistent transformation of functions to the 5-member relational network.

Control Participants

All of the control participants (C1, C2, C3 and C4) were first trained to press the spacebar at a steady rate to stimulus C, which was followed by a test for the transformation of functions involving the other four members of the relational network. The number of trials required by the control participants to meet training criterion ranged between 3 and 4 (M = 3.25, SD = .50). Three of the control participants (C2, C3 and C4) pressed the spacebar an equal number of times to each of the five test stimuli when they were presented (see Figure 5.4). C1 on the other hand, only responded when stimulus C was presented. That is, every time stimulus C was presented, C1 pressed the spacebar 30 times, but made no bar presses when the other members of the relational network were presented (see Figure 5.4). Thus, all of the control participants, failed to demonstrate the transformation of "More-than" and "Less-than" discriminative functions to the 5-member relational network.

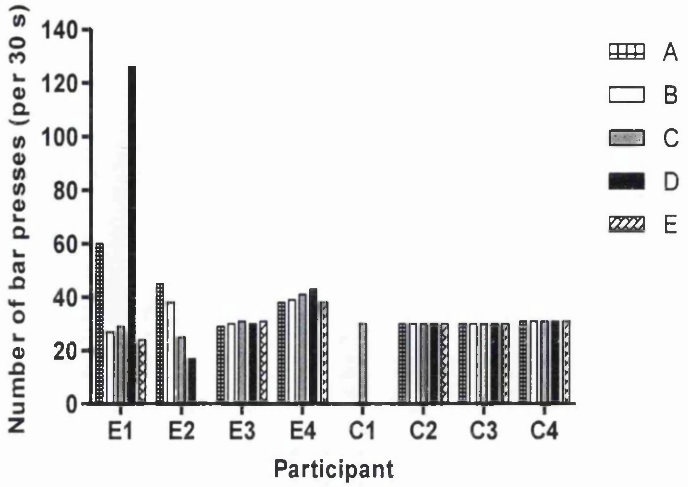


Figure 5.4. Experiment 9C. The number of bar presses per 30 s to stimuli A, B, C, D and E for the four experimental (E1, E2, E3 and E4) and control (C1, C2, C3 and C4) participants in the bar press testing phase of Experiment 9C.

In summary, the fixed order stimulus presentation during transformation tests (i.e., A-B-C-D-E) did not lead to the predicted patterns of performance for the four experimental participants in Experiment 9C. For example, E2 pressed the spacebar "more" to stimuli A and B, and "less" to stimuli D and E, than to stimulus C, thus demonstrating the transformation of functions in the opposite direction. The remaining participants (E1, E4 and E5) also failed to demonstrate consistent transformation to the 5-member arbitrary relational network. In addition, four control participants were not exposed to the non-arbitrary and arbitrary relational pre-training and testing, and also failed to demonstrate the predicted patterns of performance.

The fact that the control participants in the current experiment did not respond in accordance with the 5-member relational network during transformation tests highlights the fact that the relational training procedures were partly responsible for the performances displayed by the experimental participants. However, the failure of these participants to display the predicted patterns of performances warrants further investigation. For example, it may have been beneficial to re-expose participants to additional training and test phases if they initially failed to display the predicted behavioural patterns. Indeed, performances for a number of participants in Experiments 9A and 9B were seen to benefit from additional presentations of the test stimuli. That is, for a number of participants, performances improved from their first to last exposure to the test stimuli. Therefore, Experiment 9D aimed to examine the effects of repeated exposure to transformation training and testing.

Experiment 9D

Experiment 9D again sought to examine the potential facilitative effects of presenting the five members of the relational network in a fixed order during transformation tests. However, in comparison to Experiment 9C, all participants (experimental and control) were exposed to the bar press training and testing phase for a total of two times.

Method

Participants

Eight participants (4 experiment and 4 control), three male and five female, ranging in age from 21 to 32 years (M_{age} = 26.50, *SD* = 1.23), were recruited through personal contacts and the psychology subject pool at Swansea University. Participants received either partial course credit or £5 on completion of the task.

Procedure

The general procedure was identical to Experiment 9C, with the exception that during Phase 6, participants were exposed to the bar press training and test phases for a total of two times. Training and test mastery criterion for all phases remained the same as Experiments 9A-9C.

Results and Discussion

Of the four experimental participants that took part in Experiment 9D, all four (E1, E2, E3 and E4) successfully completed both arbitrary relational test phases, and required between 1 and 3 (*M* = 1.5, *SD* = 1.00; see Table 5.4) exposures to testing to do so. In addition, four control participants (C1, C2, C3 and C4) took part in Experiment 10D.

Table 5.4 displays the trials to criterion for participants during both constructed-response non-arbitrary relational training phases and constructed-response arbitrary relational training. Participants that passed the experimental task only required between 10 and 21 (CR-NARB1: *M* = 14.00, *SD* = 4.24; CR-NARB 2: *M* = 10.00, *SD* = 0.00) trials to reach criterion during the constructed-response non-arbitrary relational training phases. The number of trials required, to achieve constructed-response arbitrary relational training criterion, ranged between 12 and 24 ($M = 20.00$, $SD = 6.19$).

A summary of performance accuracy during Test 1 (mutual entailment) and Test 2 (combinatorial entailment) for participants that passed the experimental task can be seen in Table 5.4. During the bar press training part of phase 6 in Experiment 9D, all experimental participants required only 3 ($M = 3.00$, $SD = .00$) trials to reach criterion.

Of the four experimental participants that were exposed to tests for the transformation of "More-than" and "Less-than" discriminative functions, three (E1, E3 and E4) responded in accordance with the 5-member relational network.

Table 5.4
Trials to criterion for experimental participants during constructed-response non-arbitrary and arbitrary relational training in Experiment 9D. Also displayed is individual data for the four experimental participants.

	Phases 1A&2A: CR-NARB Relational		Phase 3: CR-Arbitrary Relational	Phase 4: Arbitrary Relational Test 1 (mutual entailment)		Phase 5: Arbitrary Relational Test 2 (combinatorial entailment)		
Participant	Training		Training	B	ME	B	CE1	CE2
1 (All-Less)	16	10	24	16/16	16/16	16/16	24/24	16/16
2 (All-More)	15	10	24	16/16	16/16	16/16	24/24	16/16
3 (All-More)	21	10	24	15/16	15/16	16/16	24/24	16/16
4 (All-Less)	12	10	24	16/16	16/16	14/16	12/24	3/16
	10	10	12	16/16	16/16	13/16	15/24	8/16
	10	10	12	16/16	16/16	16/16	24/24	16/16

Note. CR = Constructed-response and NARB = Non-arbitrary. "B" (Baseline) refers to test trials involving directly trained relations; ME = mutually entailed; and CE1 and CE2 = one- and two-node combinatorially entailed relations, respectively. Data are displayed for the number of correct responses to the baseline and mutually entailed relations during Test 1, and also to the baseline and one- and two-node relations during Test 2.

E1, E3 and E4

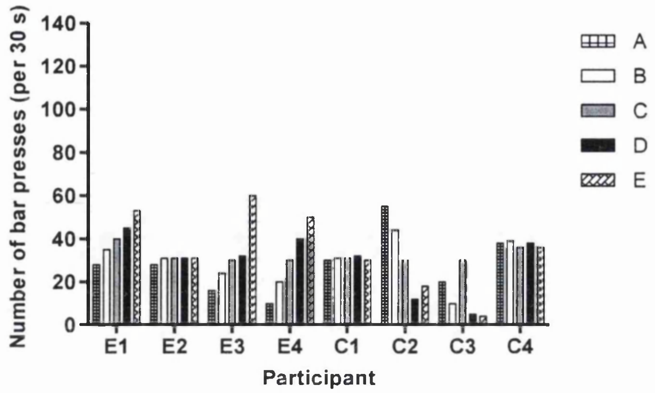
Across both exposures to the transformation test phase, E1, E3 and E4 demonstrated consistent transformation to all test stimuli. That is, these three participants pressed the spacebar "more" to stimuli D and E and "less" to stimuli A and B than to stimulus C (see Figure 5.5).

Participants who failed to demonstrate the transformation of discriminative functions

E2. On both of his/her exposures to transformation tests, E2 pressed the spacebar the same number of times to stimuli B, C, D and E, and "less" times to stimulus A. Thus, E2 failed to demonstrate the transformation of discriminative functions to the 5-member relational network.

Control Participants

Similar to Experiment 9C, all of the control participants (C1, C2, C3 and C4) were first trained to press the spacebar at a steady rate to stimulus C, which was followed by a test for the transformation of functions involving the other four members of the relational network. All of the control participants were exposed to training and testing for a total of two times. The number of trials required by all control participants to meet training criterion was 3 (M = 3.00, SD = .00). During both exposures to the transformation test phase, C1 pressed the spacebar almost an equal number of times to each of the five test stimuli (see Figure 5.5). C2, across both exposures to the test phase, pressed the spacebar "more" to stimuli A and B, and "less" to stimuli D and E, than to stimulus C. On the other hand, C3 pressed "less" to stimuli A, B, C and D, across his/her two exposures to transformation testing. Finally, on his/her first exposure to testing, C4 pressed "more" to stimuli A, B and D, than to stimulus C, and he/she pressed the same number of times to stimuli C and E. On his/her second exposure to the transformation test phase, C4 pressed "more" to stimulus D and "less" to stimulus A, than to stimulus C. C4 pressed the spacebar the same number of times to stimuli B, C and D.



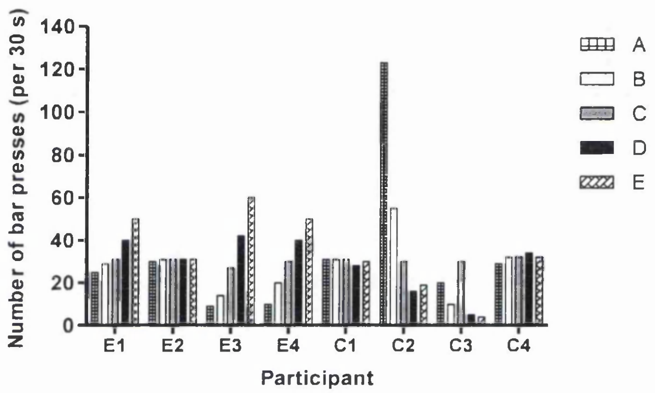


Figure 5.5. Experiment 9D. The number of bar presses per 30 s to stimuli A, B, C, D and E for the four experimental (E1, E2, E3 and E4) and control participants (C1, C2, C3 and C4) in the bar press testing phase of Experiment 9D. The upper figure refers to participants' responses during their first exposure to the transformation test phase, while the lower figure refers to responses during their second exposure to testing.

In summary, the results of Experiment 9D demonstrated that three out of four of the experimental participants (E1, E3 and E4) successfully displayed the predicted patterns of performance. In addition, these three participants demonstrated the transformation of functions across both exposures to the test phase. In comparison, none of the control participants responded in accordance with the 5-member relational network. Thus, the current findings seem to suggest that repeated exposure to training and testing may facilitate the emergence of the transformation of "More-than" and "Less-than" discriminative functions. However, further research is needed to determine the amount of training and testing required, as one participant (E2) still failed to display the predicted pattern of performance when exposed to additional training and testing phases.

General Discussion

The current set of experiments were concerned with examining the transformation of "More-than" and "Less-than" discriminative functions in accordance with a 5-member relational network (A-B-C-D-E). In addition, the current study aimed to replicate and extend Dougher et al.'s (2007) findings from a 3- to 5-member relational network, and also, address a limitation with their relational training procedures.

Experiments 9A and 9B

The purpose of Experiments 9A and 9B were to examine the transformation of discriminative functions when the test stimuli were presented in a quasi-random order, three times each during testing. This is in contrast to Dougher et al.'s (2007) study in which the test stimuli were presented only once each, in a fixed order (e.g., A-B-C), and stability of responding was not assessed. Although Dougher et al. argue that they incorporated this method of stimulus presentation on the basis of guidelines outlined for shock intensities during the respondent conditioning phase of the experiment, it was not essential that in the bar press testing phase of the experiment, participants were exposed to only one presentation of each of the test stimuli. Indeed, a previous study by Whelan et al. (2006), which examined the transformation of consequential functions in accordance with the relational frame of comparison, involved participants being exposed to each test pair, twice, in a random order, during transformation tests. With respect to the current findings, both Experiments 9A and 9B demonstrated that half of the participants responded in accordance with the 5-member relational network (i.e., press "less" to A and B and "more" to D and E, than to C). These findings do not replicate Dougher et al.'s (2007), in which all participants in Experiment 1 (8 out of 8), and five out of six participants in Experiment 2, responded "less" to the stimuli ranked lower in the network and "more" to the stimuli ranked higher in the network. Thus, exposing participants to a greater number of stimulus presentations in a quasi-random order, during transformation tests, did not lead to the predicted patterns of performance.

One potential reason for this, and a limitation to the current study, is that the immediate transfer or, transformation of functions was not assessed. For instance, immediate transfer is the concept of reaching a pre-determined mastery criterion on the first exposure to a testing block (Dymond & Rehfeldt, 2000). However, participants are often exposed to additional training and testing phases if this pattern of behaviour initially fails to emerge (e.g., Barnes & Keenan, 1993; Gil, Luciano, Ruiz, & Valdivia-Salas, 2012). One recent study to undertake such an investigation was conducted by Gil et al. (2012), in which the authors examined the transformation of functions through hierarchical relations. A mastery criterion of 6 out of 7 correct responses was incorporated during transformation testing, which, if not met, resulted in re-exposure to training and testing of the stimulus functions. The authors found that for the five participants that initially failed to pass transformation tests, re-training allowed four participants to pass subsequent transformation test phases. In turn, this method of testing may have been beneficial for the participants across all experiments in the current study that failed to demonstrate the transformation of discriminative functions. Therefore, future studies should seek to incorporate pre-determined mastery criterion during transformation tests.

Experiments 9C and 9D

The purpose of Experiments 9C and 9D were to address the fact that only half of the participants in Experiments 9A and 9B displayed the predicted patterns of performance during transformation tests. In addition, both experiments were concerned with whether the fixed order of stimulus presentation employed in the Dougher et al. (2007) study, influenced performances at test. Thus, in Experiments 9C and 9D, participants were presented with the test stimuli, once each, in the following fixed order: A, B, C, D and E. Furthermore, in Experiment 9D, all participants were exposed to transformation training and testing for a total of two times, irrespective of their initial test performances. Findings revealed that none of the participants in Experiment 9C responded in accordance with the 5-member relational network, whereas three out of four participants in Experiment 9D did.

The findings of Experiments 9C and 9D, together with those from Experiments 9A and 9B, provide some conflicting evidence regarding the emergence of the transformation of "More-than" and "Less-than" discriminative functions. For example, only half of the participants in Experiments 9A and 9B displayed responding consistent with the pre-established relational network, when the test stimuli were presented in a quasi-random order during transformation tests. In addition, none of the participants in Experiment 9C demonstrated the predicted patterns of performance when the test stimuli were presented in a fixed stimulus order, whereas, three out of four participants in Experiment 9D did so, when exposed to transformation training and testing for a total of two times. Thus, it is still unclear as to whether the fixed order of stimulus presentation, or the additional exposure to transformation training and testing phases were responsible for the successful performances observed in Experiment 9D. Furthermore, the current study examined the transformation of discriminative functions to a 5-member relational network, in contrast to the 3-member network employed in the Dougher et al. (2007). Thus, it remains to be seen whether different factors facilitate the transformation of discriminative functions to a 3- and 5-member arbitrary relational network. Further empirical work is warranted on this issue.

Relational training procedures

The current set of experiments also aimed to address an inconsistency, with the relational training and testing protocol employed by Dougher et al. For example, it is questioned as to whether participants in Experiment 1 of the Dougher et al. (2007) study were responding in accordance with the *arbitrary* relation between the comparison stimuli and the sample stimulus (contextual cue). Indeed, this apparent inconsistency was noted by Stewart and McElwee (2009) who stated that the relational training and testing procedures outlined by Dougher et al. (2007) were in fact non-arbitrary, and not arbitrary. For instance, in Experiment 1 of the Dougher et al. (2007) study, participants were presented with one symbol at the top of the computer screen, and three comparison stimuli at the bottom of the computer screen. The comparison stimuli consisted of three symbols that differed in terms of size (e.g., small, medium and large). According to Dougher et al. (2007), their training procedures established arbitrary comparative responding between the samples and comparison stimuli, however, when one considers the definition of *arbitrary* with respect to relational responding, this does not appear to be true. For instance, Stewart and McElwee (2009) point out that the relational training procedures employed by Dougher et al. used an abstract image as a contextual cue to control responding to the comparison stimulus, and not as a sample stimulus involved in a relation with the comparison stimulus. Thus, the procedures employed by Dougher et al. (2007) were used to generate non-arbitrary comparative responding under the control of abstract images (e.g., A, B and C), which later came to control responding to novel non-arbitrary stimulus sets during testing.

In turn, the procedures used by Dougher et al., may be seen as similar to the non-arbitrary relational training and testing phases employed in the current study. Therefore, the current study appears to have overcome some of the potential limitations with Dougher et al.'s training procedures, and represents a first demonstration of the transformation of discriminative functions in accordance with a 5-member relational network. In addition, the current findings extend those examining the transformation of discriminative functions from equivalence (e.g., Dymond, Whelan, & Smeets, 2005; Roche et al., 2000), to comparative relations.

However, an issue withstanding from the current experimental work, is that it may be necessary for future studies to examine the respondent and eliciting functions associated with stimuli from a 5-member relational network, using procedures similar to those employed by Dougher et al. (2007). For example, in the Dougher et al. study, following the bar press training and testing phase, participants were exposed to a respondent conditioning phase in which the middle ranking stimulus B, was paired with a mild electric shock. Testing then involved the presentation of stimuli A and C, and changes in skin conductance were recorded as the dependent measure. Findings from Experiments 1 in the Dougher et al. study demonstrated that, participants displayed higher changes in skin conductance to stimulus C and lower levels to stimulus A, than to stimulus B, even though they had never directly experienced shock associated with these stimuli. Thus, Dougher et al. (2007) propose that the behavioural processes involved in the transformation of functions may provide an alternative account to the proposed cognitive models of the clinical symptoms observed in anxiety and fear reactions (see also Roche et al., 2000; Dymond, Schlund, Roche, Whelan, Richards, & Davies, 2011). That is, the current behavioural account may have the potential to account for how individuals come to arbitrarily relate symbols and events in their environments, and thus, engage in certain avoidant behavioural patterns, even though they have never directly received reinforcement for doing so. However, if the current approach is to provide a viable alternative to cognitive models of clinically significant behaviours, then further research from an RFT perspective is warranted. In addition, it is necessary for such procedures to overcome potential confounds, such as those noted in the definition of arbitrary and non-arbitrary in the Dougher et al. (2007) study, if the current account is to accurately model how individuals come to arbitrarily relate symbols and events in their environments.

Limitations

A potential criticism of the current study centres on the length of the arbitrary relational test phases that participants were exposed to before transformation training and testing. For example, across all experiments, participants were exposed to two test phases, in which probes for the properties of mutual entailment were presented first, followed by probes for one- and two-node combinatorial entailment. In addition, during these test phases, participants were required to make a minimum of 3 out of 4 (i.e., 75% accuracy) correct responses on all test trials. Furthermore, if this criterion was not met initially, participants were re-exposed to the entire experimental task up to three further times. A potential problem with both the high mastery criterion and additional training and testing phases is that this may have, inadvertently, affected performances during tests for the transformation of discriminative functions. Indeed, the current method of presenting test blocks involving mutually entailed relations before probes for combinatorial entailment was previously employed to examine the prerequisites necessary for the emergence of relational reasoning abilities (transitive inference) in adult participants. Therefore, although it is necessary that accurate responding to the 5-member relational network is firmly established before participants are exposed to transformation tests, it may be beneficial for future studies to present both mutually and combinatorially entailed relations within the same test block, where participants are required to meet an averaged mastery criterion across all test relations. This in turn may help to circumvent potential problems associated with fatigue and inattention, which may affect performances during transformation tests.

To conclude, the current set of experiments demonstrated that a number of participants were successful in demonstrating the transformation of discriminative functions in accordance with a 5-member relational network. In addition, the current findings have overcome a potential limitation with the Dougher et al. (2007) study, and extended the examination of this pattern of responding from a 3- to 5-member network, and from a non-arbitrary, to an arbitrary stimulus set. However, as there was considerable variation in participant responding across all experiments, further research is needed to determine the factors affecting the emergence of this behaviour. In addition, it may also be beneficial for future studies to examine the respondent and eliciting functions associated with stimuli from the 5-member relational network employed in the current study, using procedures similar to those employed by Dougher et al. (2007).

Chapter 6

General Discussion

The current thesis sought to undertake an examination of the emergence of TI in typically developing adults. More specifically, the current experimental work sought to determine the potential utility of a novel behaviour-analytic account based on the principles of RFT, in generating this type of responding. Chapter 2 examined the more traditional way in which the TI problem has been studied, while Chapters 3 and 4 sought to determine the utility of establishing arbitrary comparative responding as an alternative account of TI. In addition, throughout Chapters 2-4, the utility of a range of interventions in facilitating the emergence of this behaviour were investigated. Chapter 5 undertook an examination of the transformation of discriminative functions, following the establishment of arbitrary comparative responding to a 5-member relational network. The following sections will summarise and discuss the results of Chapters 2-5, followed by a consideration of the differing theoretical accounts of TI, along with some suggestions for future research.

Chapter 2: Summary of results and discussion

Chapter 2 sought to provide a precursor to later chapters by examining the emergence of TI in adult participants. In addition, Chapter 2 sought to determine the effectiveness of a number of variables, namely, training to criterion, test mastery criterion, repeated exposures to training and test phases, and awareness, on the emergence of this behaviour. The training and testing protocols employed throughout Experiments 1-3, were similar to those currently employed by researchers examining this phenomenon in the laboratory (e.g., Frank et al., 2005; Greene et al., 2001, 2006; Lazareva & Wasserman, 2010; Moses et al., 2006, 2008; Smith & Squire, 2005; Ryan et al., 2009). However, a critical difference between the training and testing protocol employed in Chapter 2 and those currently employed, was that a pre-determined test mastery criterion was in force across Experiments 1-3. That is, participants were required to achieve a minimum mean of 80% correct across all test relations, in order to complete the experimental task. Failure to do so, resulted, in re-exposure to training and test phases, for a pre-determined number of times (e.g., 4).

Results from Experiments 1-3 demonstrated that for a number of participants, repeated exposure to training and test phases facilitated more accurate performances at test. That is, for a number of participants, the predicted patterns of performance emerged gradually, and following additional exposure to training and test phases. However, it must also be noted that, across Experiments 1-3, repeated exposure to training and testing did not facilitate the emergence of inferential responding for a number of participants. As mentioned, a further aim of Chapter 2 was to examine the role of awareness in facilitating the emergence of inferential responding in adult participants. Findings from Experiments 2 and revealed that performances were comparable between the Informed and Uninformed groups on the baseline pairs, but more accurate performances were observed for the Informed group on the endpoint and non-endpoint pairs. More specifically, in Experiment 2, accuracy on the critical BD non-endpoint pair was only just above chance levels for the Uninformed group (54%), while accuracy on the non-endpoint pair, BE, was slightly higher (63%). In contrast, accuracy for the Informed group in Experiment 2 on these same test pairs was considerably higher (Informed: BD: 81%; BE: 76%). Comparable, high levels of accuracy were observed between the groups on the non-endpoint pair CE. Similar findings were observed in Experiment 3. However, in comparison to Experiment 2, participants in the Uninformed group performed above chance levels on the non-endpoint pairs, BD (64%) and BE (72%). Furthermore, Experiment 3 found that post-experimental measures of awareness for both the Informed and Uninformed groups were not correlated with performances on the non-endpoint, inferential pairs at testing.

As mentioned, a primary aim of Experiment 1 was to explore the utility of incorporating test mastery criterion and repeated exposure to training and testing phases, in generating inferential responding. Results from Experiment 1 demonstrated that a number of participants successfully met criterion at testing, and achieved high levels of accuracy on the baseline, endpoint and non-endpoint pairs. However, it must also be noted that, a number of participants in Experiment 1 failed to meet criterion at testing, and repeated exposure to training and test phases, did not allow accurate performances to emerge. The incorporation of additional training and test phases in Experiment 1 is in contrast to the method in which inferential responding is currently examined, in that participants are only provided with one opportunity to respond to test trials (e.g., Acuna et al., 2002; Frank et al., 2005; Lazareva & Wasserman, 2010; Moses et al., 2006, 2008). In comparison, if weak inferential performances are initially observed, then re-exposure to training and testing phases may lead to improvements in performance accuracy on inferential probe trials. However, in order to more clearly determine the potential facilitative effects of repeated exposures to training and testing on the emergence of TI, it may be necessary for future studies to include control conditions, in which participants are not exposed to additional training and test phases.

In addition, Experiments 1-3 incorporated a test mastery criterion of 80%, which allowed us to examine the emergence of stable patterns of predicted and unpredicted performances (Dymond & Rehfeldt, 2000). For example, across all experiments in Chapter 2, participants were required to achieve a minimum mean of 80% across all fifteen, test pairs, in order to pass the experimental task. Findings from Chapter 2 revealed that there were yields of 44% in Experiment 1, 70% in Experiment 2, and 72.5% in Experiment 3. Furthermore, findings from Chapter 2 also highlighted that a number of participants across all experiments were unable to meet criterion at testing. In addition, an interesting finding from all experiments in Chapter 2 was that despite the implementation of mastery criterion during test phases in Experiments 1-3, a number of participants were capable of passing the experimental task, in the absence of inferential responding. That is, a number of participants demonstrated criterion performances on the baseline and endpoint test pairs, but not on the non-endpoint pairs. In turn, such findings seem important for researchers seeking to determine whether inferential responding has emerged, and whether cross-species generalisations about the strategies employed to solve the task, can be made. For instance, currently in studies examining TI, there is no standard accuracy criterion that participants must achieve in order to demonstrate TI. However, if a pre-determined mastery criterion was employed during test phases, this may provide researchers with a more reliable method of determining whether successful inferential responding, has, in fact, emerged. Future studies should therefore seek to incorporate mastery criterion during test phases in order to explore this issue.

Findings from both Experiments 2 and 3 revealed that awareness of the stimulus hierarchy led to more accurate responses at test. For example, in both Experiments 2 and 3, participants in the Informed group displayed more accurate responses to the endpoint and non-endpoint test pairs, in comparison to participants in the Uninformed group. However, in Experiment 3, improvements were noted for participants in the Uninformed group on the non-endpoint pairs BD and BE (BD: 64% and BE: 72%), over those noted in Experiment 2 (BD: 54% and BE: 63%). Findings from Experiments 2 and 3 in Chapter 3 are important for a number of reasons. For example, the findings from Chapter 2 may contribute to the current debate regarding the number of strategies humans employ to solve the task. Some researchers argue that, if individuals rely on conscious awareness, then multiple strategies are employed during learning of the task, and indeed other relational learning tasks (e.g., Martin & Alsop, 2004; Moses et al., 2006, 2008; for a review, see Willingham, 1997). In turn, these findings are important if cross-species generalisations are to be made about the strategies that humans and non-humans employ to solve the task (e.g., Moses et al., 2006, 2008). However, it is difficult to definitively conclude the specific role of awareness on the emergence of TI, due to the method in which it was assessed throughout Chapter 2. For example, and as mentioned in Chapter 1, some researchers propose that there is strong evidence that post-experimental measures of awareness underestimate contingency awareness (e.g., Dawson & Reardon, 1973). Indeed, findings from Experiment 3 revealed that providing participants with additional instructions at the start of the experiment did not guarantee awareness at the end of the experiment. More specifically, accurate performances on the inferential, non-endpoint pairs, BD, BE and CE for participants in the Informed group, were not correlated with post-experimental measures of awareness. In addition, the disparity observed across studies examining the role of awareness when similar training and testing protocols are employed highlights the fact that more reliable methods of measurement are needed. Therefore, if concurrent measures were taken whilst participants were engaged in the task, this may help to provide a more robust measure of awareness (see Lovibond & Shanks, 2002). Future studies should seek to undertake such an investigation.

A further issue arising from Chapter 2 is the fact that the predicted patterns of performance did not emerge immediately for a number of participants. This in turn poses some challenges to cognitive accounts of TI. For example, the Image theory suggests that individuals possess an innate capacity to construct mental models of the stimuli (e.g., Johnson-Laird, 1972). More specifically, the Image theory suggests that all baseline pairs are integrated into a mental line during training, which allows individuals to solve inferential problems at test by conducting a spatial search of this information (de Soto et al., 1965; Huttenlocher, 1968). However, findings from

Chapter 2 would suggest otherwise. This is seen in the fact that for a number of participants across Experiments 1-3, accuracy was not perfect on all test pairs, inferential responding did not emerge immediately, nor did it emerge after repeated exposures to training and testing. Therefore, if participants in Chapter 2 were merely required to consult the mental line to solve inferential problems at test, then why were a number of them unable to do so? Furthermore, the Image theory does not include any suggestions as to why this behaviour fails to emerge, nor does it include any suggestions on how to remediate such weaknesses (e.g., Vasconcelos, 2008). In contrast, these are very important issues inside the realm of behaviour analysis, as evidenced throughout Experiments 1-3 in Chapter 2. In addition, the development and implementation of a range of interventions to generate accurate responding and remediate weaknesses are an important goal in behaviour-analytic research (for a review, see Vitale et al., 2008), and thus, Chapters 3 and 4 of the current thesis, further explored this issue.

In summary, findings from Chapter 2 highlighted the potential utility of repeated exposure to training and test phases in facilitating the emergence of inferential responding in adult participants. In addition, findings from Experiments 2 and 3 propose that awareness of the stimulus hierarchy leads to more accurate performances on the TI task. However, findings from Experiment 3 revealed that participants in the Uninformed group performed above chance levels on all non-endpoint, inferential pairs, and that providing participants with additional instructions at the start of the experiment, does not guarantee awareness at the end of the experiment. The implications of these findings will now be discussed.

Implications of findings

The findings from Experiments 2 and 3 regarding the role of awareness in the emergence of TI, may have important implications for our understanding of the emergence of this pattern of behaviour in individuals that suffer from cognitive impairments, such as amnesia (e.g., Smith & Squire, 2005) and schizophrenia (e.g., Armstrong, Kose, Williams, Woolard, & Heckers, 2010; Coleman, Titone, Krastoshevsky, Krause, Huang, Mendell, Eichenbaum, & Levy, 2010; Titone et al., 2004). For example, a number of studies have found that the ability to make inferential judgments is impaired in individuals with a diagnosis of schizophrenia (e.g., Titone et al., 2004). More specifically, Titone et al. (2004) compared performances on the TI task for a group of schizophrenia patients and a group of healthy controls, and found that performance on the BD inferential pair was impaired for the schizophrenia patients (see also Coleman et al., 2010). However, it must be noted that the schizophrenia patients were able to learn the training pairs, and display high levels of accuracy on the endpoint pairing AE, at comparable levels to the controls. Coleman et al. (2010) propose, that, such findings reflect an inability of these individuals to manipulate the stimulus hierarchy in order to make a correct response. In addition, Coleman et al. (2010) found that conscious awareness was not significantly associated with the ability to make an inferential judgment for schizophrenia patients. Thus, Titone et al. (2004) suggest that awareness is not the source affecting the emergence of TI in schizophrenia. In addition, findings from a study by Armstrong, Williams and Heckers (2012) support both of the aforementioned studies, and propose that individuals with a diagnosis of schizophrenia suffer from a differential relational memory deficit, and that awareness is not the contributing factor. However, in this study, the authors investigated the potential facilitative effects of exposing patients with schizophrenia to reduced-sized training blocks and additional feedback during training, in an attempt to maximise the number of patients that were exposed to inferential tests. Armstrong et al. (2012) found that in comparison to a previous study (e.g., Armstrong et al., 2010), only 8% (3/37) of the schizophrenia patients, failed to complete training. That is, reduced-sized training blocks and feedback during training, allowed a greater number of patients with schizophrenia to be exposed to inferential test phases. The training and testing protocol employed throughout Experiments 1-3 in Chapter 2 may therefore have the potential to provide an alternative method to examine the emergence of TI in individuals suffering from cognitive impairments, such as schizophrenia. Furthermore, if additional feedback and reduced training blocks were incorporated alongside additional training and test phases, it may be possible to generate more accurate inferential responding in patients with schizophrenia. Future studies should therefore seek to undertake such an investigation.

Chapter 3: Summary of results and discussion

Chapter 3 was the first empirical chapter to examine the potential utility of derived comparative responding as a novel account of TI-like responding in adult humans. This work also attempted to replicate and extend previous findings, which have found the account has the potential to generate derived comparative responding in adults and children (e.g., Gorham et al., 2009; Munnelly et al., 2010; O'Hora et al., 2002; Reilly et al., 2005; Whelan et al., 2006). In contrast to Chapter 2, participants throughout all experiments in Chapter 3 were exposed to training and testing on a number of conditional discriminations. With respect to Chapter 3, participants were exposed to two training and test phases, non-arbitrary (i.e., with stimulus sets that were physically similar) and arbitrary (i.e., with stimulus sets that were physically dissimilar) relational training and testing. Thus, across all phases in Experiments 4-5B in Chapter 3, and in comparison to the training and testing protocol employed in Chapter 2, participant selections were made on the basis of the contextual cue presented, and not the association between the two stimuli presented onscreen. In addition, Chapter 3 sought to explore whether some of the characteristic features noted in studies of TI, would be noted with a Relational frame interpretation, based on derived comparative responding (see Munnelly et al., 2010; Reilly et al., 2005).

Findings from Experiment 4 revealed that six out of ten participants in the All-More group and nine of out ten participants in the All-Less group, met criterion at testing, and displayed high levels of accuracy on all test relations (baseline, mutually entailed, and one- and two-node combinatorial entailment). However, it was noted that for participants from both the All-More and All-Less groups that failed to meet criterion during arbitrary relational test phase, accuracy was low on the mutually entailed relations. Thus, it was proposed that the contextual functions of MORE-THAN and LESS-THAN established during the non-arbitrary relational phases, were not functioning during the arbitrary relational phases. That is, if participants selected B not A (B+A-) in the presence of the MORE-THAN cue during arbitrary relational training, then they also selected B not A in the presence of the LESS-THAN contextual cue. Experiments 5A and 5B were designed in an attempt to address this issue, and both experiments incorporated an observing response and a variant of the simple-to-complex protocol, in an attempt to generate greater stimulus control.

Findings from Experiments 5A and 5B revealed that five out of ten participants in Experiment 5A passed the experimental task, while only one out of eight participants in Experiment 5B, met criterion at testing. Thus, the addition of the observing response and the variant of the simple-to-complex protocol in Experiments 5A and 5B resulted in only limited improvements in stimulus control.

The results of Experiment 4 revealed that the current model has the potential to generate arbitrary comparative responding which is comparable to TI-like behaviour in adult participants. A second aim of Experiment 4 was to examine an issue in a previous study by Munnelly et al. (2010), in which low levels of accuracy were observed for participants on the mutually entailed relations, and the one- and two-node relations that were *different to training*. Thus, Experiment 4 explored whether increasing the number of trials presented at test would lead to improvements in response accuracy. For instance, in the Munnelly et al. (2010) study, participants were only exposed to one presentation of each test pair, which made it difficult to determine whether participants were randomly responding to the test stimuli. Thus, exposing participants to two presentations of each test pair, allowed us to control for the potential confound of chance performances. For example, if participants responded correctly to both presentations of a test pair, then we may propose that effective stimulus control over responding was established. However, if participants responded incorrectly to both presentations of a test pair, we may propose that stimulus control over responding was lacking. Findings from Experiment 4 seem to have overcome some of the potential problems associated with the Munnelly et al. (2010) study, in that the average performance accuracy for participants in Experiment 4 was above chance levels on all test relations. However, upon closer inspection of the data from Experiment 4, low accuracy was observed for a number of participants that failed to meet test criterion on the mutually entailed relations, and one- and two-node relations that were *different to training*. Thus, increasing the number of test trials in Experiment 4 did not have the desired facilitative effects on performances for participants that failed to meet test criterion. Experiments 5A and 5B were therefore designed in an attempt to explore the utility of alternative interventions in generating more accurate derived comparative responding.

Experiments 5A and 5B explored the potential utility of a variant of the simple-to-complex protocol in facilitating the emergence of derived comparative responding. In addition, an observing response was incorporated during arbitrary relational training and testing phases, whereby the contextual cue appeared first onscreen, and participants were required to press the spacebar to allow the comparison stimuli to appear. Findings from Experiment 5A revealed that the variant of the simple-to-complex protocol has the potential to generate derived comparative responding for a number of participants, and in doing so extend previous research studies examining the effectiveness of this protocol, from equivalence (e.g., Adams et al., 1993; Fields et al., 2000; Smeets et al., 2003) to derived comparative relations. However, it must also be noted, that despite the implementation of a variant of the simple-to-complex and the observing response interventions, responding was low on the mutually entailed relations for a number of participants. Results from Experiment 5B, failed to replicate those of Experiment 5A. Although the reasons for this remain unclear, the primary difference between Experiment 5A and 5B was that participants were not re-exposed to non-arbitrary relational training and testing, if the predicted patterns of performance failed to emerge immediately. In contrast, a number of studies have found evidence for the facilitative effects associated with exposure to non-arbitrary relational training and testing when arbitrary comparative relations were found to be weak (e.g., Berens & Hayes, 2007; Gorham et al., 2009; Vitale et al., 2008). Indeed, RFT states that a history of responding in accordance with contextually controlled non-arbitrary comparative relations is an important precursor that facilitates the development of arbitrarily applicable derived relational responding (e.g., Berens & Hayes, 2007; Hayes et al., 2001a; Stewart & McElwee, 2009). Thus, a limitation to Experiment 5B was that participants were not re-exposed to non-arbitrary relational training and test phases when these weaknesses were noted, and thus, future studies should seek to take this into consideration.

One issue arising from the results across Experiments 4-5B in Chapter 3 was that despite the implementation of a number of interventions (e.g., repeated exposure to training and testing, observing response and a variant of the simple-to-complex protocol), some participants still failed to meet criterion at test. More specifically, weaknesses were observed in participants' ability to respond in accordance with the properties of mutual entailment, and one- and two-node combinatorially entailed relations that were *different to training*. In addition, it was widely noted that participants that failed to meet test criterion, were able to respond in accordance with the test relations that were the *same as training*. This in turn led to the question being posed as to whether the weaknesses noted, were in fact, as a result of the conditional discrimination training and testing protocol employed. For example, with the conditional discrimination, participants were presented with one of two sample stimuli (i.e., contextual cue) in the top portion of the computer screen, followed by two comparison stimuli in the bottom left and right of the screen. Thus, the conditional discrimination may represent a tendency for individuals to have a preference for constructing spatial arrays from a top-down perspective (Evans et al., 1993). However, Reilly et al. (2005) argue that, from a behaviour-analytic point of view, this argument is incomplete, and thus, a demonstration of preferences for establishing top-down stimulus arrays, would need to be verified empirically. For example, it may have been possible that as a result of the spatial arrangement of the stimuli, participants ignored the contextual cue. That is, was it possible that the contextual cue was not discriminative for relational responding, and similar to Chapter 2, one of the comparison stimuli in a learning pair became the discriminative S+ for reinforcement, while the other became the S- for non-reinforcement? If participants were in fact ignoring the contextual cue, then this may potentially account for why they were able to display high levels of accuracy on the relations that were the *same as training* in comparison to those that were *different to training*. Thus, it remains to be seen whether relations that are *different to training* are indeed more difficult to solve, or whether in fact, participants may ignore the contextual cue during arbitrary relational training and testing phases. Further empirical work is needed on this issue, and thus, Chapter 4 sought to further explore this by incorporating a variant of the RCP, in an attempt to generate more accurate responding.

In summary, the findings from Experiments 4-5B in Chapter 3 provide evidence that responding to a combination of "More-than" and "Less-than" derived comparative relations, has the potential to provide a novel behaviour-analytic account of TI-like behaviour in adult participants. However, findings from all experiments revealed that further research is needed to determine the conditions that are conducive to the emergence of this behaviour. For instance, despite the implementations of a number of interventions (repeated exposure to training and test phases, an observing response, and a variant of the simple-to-complex protocol), responding on the mutually entailed relations was weak, for a number of participants. In addition, it was proposed that the conditional discrimination may have contributed to these failures in stimulus control, and thus, Chapter 4 sought to explore this issue by incorporating a novel procedure, the RCP, to examine arbitrary comparative responding. The implications of findings from Chapter 3 will now be discussed.

Implications of findings

The conditional discrimination employed to examine TI-like behaviour in adult participants may bear relevance to an examination of the brain regions recruited during performances on the TI task. For example, a number of studies have proposed that the ability to solve the task is mediated by an interaction between several neural systems (e.g., Dusek & Eichenbaum, 1997; Moses et al., 2006, 2008; Waltz, Knowlton, Holyoak, Boon, Mishkin, deMeneyes Santos, et al., 1999). Findings from lesion studies have demonstrated that the frontal and medial temporal lobes are recruited during the TI task (e.g., Dusek & Eichenbaum, 1997; Smith & Squire, 2005; Waltz et al., 1999). These proposals are further supported by neuroimaging studies, which have also implicated a role for the parietal lobe (e.g., Acuna et al., 2002; Greene et al., 2006). In addition, hippocampal activation (e.g., Greene et al., 2006; Heckers et al., 2004; Nagode & Pardo, 2001) and activation of the basal ganglia (Frank, Seeberger, & O'Reilly, 2004; Frank et al., 2005) have also been noted during investigations of TI. Thus, the current model may be well suited to such investigations. Indeed, Hinton et al. (2010) incorporated the protocol employed in Experiment 4 of Chapter 3, to examine the neural correlates of transitive inference. The authors reported a number of important findings, the first of which relates to the SDE. For example, Hinton et al. reported a clear correspondence between the SDE and activation across the parietal and prefrontal cortex. That is, the greatest brain activation was noted during the more difficult, adjacent test pairs (e.g., B>A and A<B), while the least activation was noted with the non-adjacent, two-node relations (D>A and A<D). Secondly, the authors reported that greater activation was noted in these same brain regions when participants were exposed to test pairs, devoid of end items (e.g., D>B and B<D),

when compared to those with end items (e.g., A<D and D>A). Finally, the authors reported that test trials that were *different to training*, required greater parietal activity and longer reaction times, than those that were the *same as training*. Therefore, findings from the Hinton et al. (2010) study provide further support for the utility of the conditional discrimination employed throughout Chapter 3, in providing an alternative account of performances of the TI task. In doing so, these findings also suggest that the model has the potential to be incorporated in future studies examining the neural correlates of TI.

Chapter 4: Summary of results and discussion

Chapter 4 aimed to extend the work of Chapter 3 and address a number of issues that arose in the previous chapter. For example, findings from a number of experiments in Chapter 3 revealed that participants encountered difficulties in demonstrating the targeted relations. Therefore, across all experiments in Chapter 4, the utility of a variant of the RCP in generating arbitrarily applicable comparative responding was investigated. Furthermore, the potential utility of a constructed-response protocol was examined throughout all experiments in Chapter 4. With this protocol, participants were required to construct "relational sentences" in the upper portion of the screen. Therefore, Chapter 4 sought to determine whether the evaluative properties of the RCP alongside the constructed-response format, would result in improvements in participant responding in comparison to Chapter 3. Finally, Chapter 4 sought to examine the effects of the linearity of arbitrary baseline relations on the emergence of derived comparative responding.

The results from Experiment 6A and 6B revealed that ten and eleven out of twelve participants, respectively, passed test for both mutual and combinatorial entailment. In addition, the linearity of the arbitrary baseline relations was not found to have an effect on the emergence of this behaviour. Experiments 7A and 7B again sought to explore the effects of linearity on the emergence of arbitrary comparative responding. In addition, both experiments sought to determine whether the predicted patterns of performance could be established employing a constructed-response format only. Findings from Experiments 7A and 7B were in contrast to the findings from Experiments 6A and 6B, and revealed that only two out of twelve participants in

Experiment 7A successfully completed the experimental task, while seven out of twelve participants in Experiment 7B did. For the participants across both experiments that failed to successfully meet test criterion, weaknesses were noted on their ability to respond in accordance with the properties of mutual and combinatorial entailment. In Experiments 8A and 8B, the utility of an MET intervention by means of exposure to additional non-arbitrary relational training and testing phases, revealed that all four participants that took part in Experiment 8A successfully completed the experimental task, while two out of the four participants in Experiment 8B completed the task.

The results from Experiments 6A and 6B appear to have overcome some of the limitations noted with the conditional discrimination in Chapter 3, and demonstrated that a large number of participants successfully completed the experimental task when the RCP was employed to establish responding in accordance with these relations. In addition, the findings of both experiments taken together suggest that the linearity of the baseline relations is not central to the emergence of arbitrarily applicable comparative responding, which is contrary to proposals among the literature on TI (e.g., Hunter, 1957; Russell et al., 1996). However, these findings do extend the literature on RFT examining the effects of linearity on derived comparative responding, from a 3- (e.g., Vitale et al., 2008) to 5-member relational network. In addition, these findings extend those examining the utility of the RCP for examining multiple stimulus relations from "Same" and "Opposite" to "More-than" and "Less-than". In turn, the RCP has the potential to examine other multiple stimulus relations, including "Before" and "After" and "Spatial".

The comparably high accurate test performances observed across Experiments 6A and 6B may be partially explained by the response requirements of the RCP. For instance, and as previously mentioned, the RCP was designed in an attempt to provide participants with a more evaluative method of responding. Thus, participants first completed the "relational sentence" in the upper portion of the screen, and then evaluate their selection via one of two confirmatory buttons (Finish Trial/Start Again) at the bottom of the screen. In comparison, the conditional discrimination employed throughout Chapter 3 and other studies examining derived comparative relations (e.g., Munnelly et al., 2010; Reilly et al., 2005), required participants to press a designated key on the left or right of computer keyboard to make their selection, and immediately upon doing so, a new trial was initiated. Thus, participants were not provided with an opportunity to evaluate their response before the next trial commenced. When Dymond and Whelan (2010) explored the evaluative component of the RCP, the authors found that it had a facilitative effect on performances in terms of overall yield when compared to the MTS. Furthermore, Dymond et al. (2013) argue that, in comparison to the MTS, discriminative control may potentially be enhanced as the selected comparison stimulus is placed on the same level as the contextual cue and sample. That is, the authors propose that moving the comparison stimulus away from the rejected stimulus, may enhance discriminative control (Dymond & Whelan, 2010). This is in contrast to the method in which feedback was presented with the conditional discrimination employed across all experiments in Chapter 3. For example, across Experiments 4-5B in Chapter 3, during non-arbitrary and arbitrary training phases, after participants made their selection, the screen cleared and only the feedback "Correct" or "Wrong" appeared onscreen. Dymond and Whelan (2010) propose that discriminative control may be diminished with the top-down presentation of stimuli in the MTS, or conditional discrimination. Indeed, with respect to the current thesis, failures in stimulus control were greatly reduced in Experiments 6A and 6B, and thus, the RCP may hold the potential to provide a viable alternative to the conditional discrimination for establishing effective stimulus control over responding to arbitrary comparative relations.

The influence exerted by the sequential presentation of stimuli from left-to-right in the current study highlights the importance of stimulus sequences in experimental tasks. Indeed, the RCP was developed to mimic the verbal relational processes involved in everyday tasks, such as reading and sentence-completion. In addition, and as previously mentioned in Chapter 4, Mackay and Fields (2009) propose that the position of events in sequence is critical for performances on learning tasks. For instance, the authors propose that, in a given language, the order of words is critical to communicating and understanding information, and this may be reflected in the subject-verb-object (S-V-O) construction noted in English language. In the following example, the syntactical ordering of words is important in our ability to effectively communicate information: "Adam sings songs". However, if the order of the words was changed, the meaning of the utterance may not make sense in English (e.g., "Adam songs sings"). Furthermore, Mackay and Fields (2009) propose that once an individual has learned a few utterances, in the S-V-O order just described, an infinite number of novel utterances may emerge without direct training. Thus, the ordering of stimuli such as that seen in the RCP may have the potential to reflect how individuals learn to produce novel utterances and sentences in their daily lives. Future research should therefore seek to undertake such an investigation with the RCP.

The findings from Experiments 7A and 7B, in which a constructed-response format was employed, were less convincing than those from Experiments 6A and 6B. Although the reasons for these findings were unclear, it was proposed that the additional exposure to non-arbitrary stimulus sets that participants in Experiments 6A and 6B encountered, may have facilitated the emergence of this pattern of responding. In addition, participants in Experiments 7A and 7B were only exposed to the experimental task in a constructed-response format, whereas those in Experiments 6A and 6B received non-arbitrary relational training and testing prior to the constructed-response protocol. Therefore, in order to isolate whether the lack of non-arbitrary training trials, or the constructed-response format contributed to the low yields in Experiments 8A and 8B, participants in Experiments 8A and 8B were exposed to a MET intervention by means of an additional non-arbitrary relational training and test phase. The constructed-response format was employed in both Experiments 8A and 8B, and findings demonstrated that the MET facilitated responding for all participants (4 out of 4) in Experiment 8A, and half of the participants (2 out of 4) in Experiment 8B. Thus, it appears that additional training with non-arbitrary stimulus sets has the potential to facilitate arbitrary comparative responding, and in doing so, extends previous findings highlighting the utility of MET interventions (e.g., Barnes-Holmes, et al., 2004; Berens & Hayes, 2007; Gorham et al., 2009; Luciano et al., 2007; Vitale et al., 2008). However, further research is needed to determine the amount of MET training that is needed to generate accurate responding as two participants in Experiment 8B, were still unsuccessful in responding in accordance with the arbitrary comparative relations of "More-than" and "Less-than".

It must also be noted that there are some potential limitations to the RCP. For example, the very fact that the RCP was developed to mimic the processes involved in everyday activities may restrict the applicability of the protocol. The RCP involves the presentation of stimuli from left-to-right, which may restrict the protocol to English-speaking countries. That is, in non-English languages such as Arabic and Hebrew, stimuli are presented from right-to-left, and thus, it is difficult to determine whether the RCP would generate similar findings in such languages. Future studies should therefore seek to undertake a comparison of performances on the RCP with populations that read from right-to-left against populations that read from left-to-right. Furthermore, and, as previously mentioned, in the English language, the subject-verb-object sentence construction is important in the communication of information. However, in languages, such as Irish, French and Spanish, possession or ownership affects the order of word placement in sentences. For instance, the Irish translation for "The school" is "An Scoil". However, when talking about the school gates, the Irish translation is "Geata na Scoile", which means "Gates of the School" in English. Thus, if the RCP were to be employed with populations in which the order or placement of words differs from the English language, then its applicability may be further limited. Future studies should seek to develop variations of the RCP for use with different languages and undertake a comparison of its utility in generating arbitrary comparative responding across languages.

Thirdly, if the RCP is to have applicability in populations with learning impairments, then a number of steps may need to be taken to ensure such populations are capable of responding to the task. For example, some individuals with learning impairments encounter difficulties on conditional discriminations (e.g., Lowenkron, 1991; Michael, 1985; Sundberg & Michael, 2001; Sundberg & Sundberg, 1990), which often involve response forms that are often more complicated than they initially appear (Sundberg & Michael, 2001). Similarly, with the RCP, participants are required to attend to a number of stimuli simultaneously, complete "relational sentences", and then confirm, or re-do that trial. Thus, the level of complexity involved in making a response with the RCP, may limit its applicability to populations with learning difficulties. However, in order to overcome some of these potential problems, it may be necessary to examine the pre-requisites necessary for responding on the task. For example, it may be beneficial to first expose participants to a familiarisation phase, in which they are assessed on their ability to respond to the drag-and-drop requirements associated with the RCP. If participants are unable to demonstrate the required response form, then it may be necessary to implement interventions to establish these repertoires. Examining the necessary response requirements in this manner may help to determine whether the individual possesses the necessary skills, to respond to, derived comparative relational tasks.

In summary, the findings from Experiments 6A and 6B revealed that the RCP has the potential to generate arbitrarily applicable comparative responding in adult participants. In addition, findings from both experiments revealed that the linearity of the training pairs was not critical to the emergence of this behaviour. In contrast to Experiments 6A and 6B, the findings of Experiments 7A and 7B failed to replicate these findings, and suggest that a history of responding to non-arbitrary relations may be an important precursor to arbitrary comparative responding. Finally, Experiments 8A and 8B attempted to address the high attrition rates from Experiments 7A and 7B by the inclusion of additional non-arbitrary training as a form of MET. For the most part, this intervention was found to have a facilitative effect on performances, and thus, replicates and extends previous findings on the utility of MET interventions. The implications of these findings will now be discussed.

Implications of findings

The RCP protocol employed throughout Chapter 4 may have important implications for future studies seeking to examine the emergence of relational reasoning abilities and transitive inference (TI) in both human and non-human populations that lack sophisticated verbal behaviour. For example, an ordinal relation may emerge between stimuli A and C (e.g., A+C-; where "+" and "-" represent the reinforced and non-reinforced responses, respectively) once a relation has been established between stimuli A and B (A+B-), and B and C (B+C-; Vasconcelos, 2008). Indeed, numerous studies have found evidence for the expression of TI in young children and non-humans when minimal, or no instructions were provided (e.g., Bryant & Trabasso, 1971; Lazareva et al., 2004; McGonigle & Chalmers, 1977). One such study, conducted by McGonigle and Chalmers (1977), involved training squirrel monkeys on the relation between four adjacent stimulus pairs (A+B-, B+C-, C+D- and D+E-), followed by tests involving non-adjacent stimulus pairs (e.g., BD). In this study, no verbal instructions were employed and the stimuli consisted of weighted cans consisting of different colours. In effect, McGonigle and Chalmers first trained the squirrel monkeys to associate "weight" with "colour", which led to successful selections of B over D in the BD pair during the critical inferential test phase (e.g., B was "lighter" than D). With respect to the current account of TI-like behaviour, it may be possible to examine the pre-requisites necessary for individuals with limited verbal behaviour to demonstrate arbitrarily applicable relational responding. For example, and as previously mentioned in Chapter 1, there is increasing evidence linking language to derived relational responding (e.g., Gross & Fox, 2009). In the current thesis, individuals with sophisticated verbal behaviour were employed as the population sample. Thus, exposing individuals with limited verbal behaviour to the RCP protocol, may allow us to identify the different forms of responding that are necessary to establish these repertoires. That is, the current protocol could first examine whether responding to non-arbitrary relations is effectively established, before examining responding to mutually, and then combinatorially entailed relations. Indeed, RFT proposes that responding in accordance with non-arbitrary relations is important in the emergence of arbitrarily applicable relational responding (e.g., Berens & Hayes, 2007; Hayes et al., 2001; Stewart & McElwee, 2009). However, to date, few studies examining arbitrary comparative responding with young children, have examined responding in accordance with non-arbitrary relations before examining arbitrary comparative responding (e.g., Barnes-Holmes et al., 2004; Gorham et al., 2009). Indeed, when Gorham et al. (2009) found that responding to arbitrary comparative relations was weak for a number of children, converting the weak arbitrary comparative relations into a non-arbitrary form resulted in improvements in response accuracy. The RCP protocol may therefore have the potential to be employed as an assessment tool, similar to the Training and Assessment of Relational Precursors and Abilities (TARPA; Moran, Stewart, McElwee, & Ming, 2010), to examine the conditions necessary to develop fluid and functional linguistic repertoires. Future studies should seek to expose populations, with language difficulties, to the RCP.

Chapter 5: Summary of results and discussion

Chapter 5 aimed to examine the transformation of "More-than" and "Less-than" discriminative functions to a 5-member relational network with adult participants. Experiments 9A-9D in Chapter 5, sought to replicate and extend Dougher et al.'s

(2007) findings from a 3- to 5-member network. With respect to Chapter 5, a function was trained to stimulus C, and testing occurred with stimuli A, B, C, D and E. On the basis of the findings from the Dougher et al. (2007) study, it was predicted that participants would respond "less" to stimuli A and B, and "more" to stimuli D and E, than to stimulus C. Chapter 5 also sought to determine the conditions necessary in facilitating the emergence of this pattern of responding. However, an important procedural difference existed between Experiments 9A and 9B in Chapter 5 of the current thesis and Experiments 1 and 2, in the Dougher et al. (2007) study. For example, in Experiments 9A and 9B of the current thesis, the five members of the relational network were presented three times each, in a quasi-random order during transformation tests. In contrast, in the Dougher et al. (2007) study, stimuli during testing were presented once each, in a fixed stimulus sequence (A-B-C). However, Experiments 9C and 9D of Chapter 5 incorporated a similar method to Dougher et al. (2007) during the transformation test phase (A-B-C-D-E).

Results from Experiments 9A and 9B revealed that, only four out of nine, and four out of eight participants, respectively, demonstrated the predicted patterns of performance, when the stimuli during transformation tests, were presented in a quasi-random order. When the test stimuli in Experiments 9C and 9D were presented in a fixed stimulus sequence, findings revealed that none of the participants in Experiment 9C demonstrated the transformation of functions, while three out of four participants in Experiment 9D, did.

The findings from Experiments 9A and 9B provide some tentative evidence for the transformation of "More-than" and "Less-than" discriminative functions to a 5-member arbitrary relational network. More specifically, the function attached to stimulus C in the 5-member relational network, transformed the function of the other members of the network, so that a number of participants responded "less" to stimuli A and B, and "more" to stimuli C and D. However, it must be noted across both Experiments 9A and 9B, only half of the participants demonstrated the predicted patterns of performance. Although the reasons for this were unclear, and as previously mentioned, it may have been possible that the method of stimulus presentation during transformation testing facilitates the emergence of this behaviour. Thus, Experiments 9C and 9D sought to explore this. Findings from Experiment 9C revealed that none of the participants exposed to transformation training and testing responded in accordance with the 5-member arbitrary relational network. In contrast, three out of four participants in Experiment 9D did. However, a critical difference between Experiments 9C and 9D, was that participants in Experiment 9D were exposed to the transformation training and test phase, for a total of two times. Potential causes of these differences will now be explored.

As a result of the findings from all experiments in Chapter 5, there are still a number of issues regarding the conditions necessary to generate the transformation of discriminative functions in adult humans that need to be addressed. For example, Dougher et al. (2007) found evidence for the transformation of "More-than" and "Less-than" discriminative functions to a 3-member comparative network, when the test stimuli were presented in a fixed sequence during transformation tests. Findings from all experiments in Chapter 5 of the current thesis do not conclusively support these findings. More specifically, the findings from Experiment 9D, and those from participants in Experiments 9A and 9B appear to suggest that, additional exposures, or an increased number stimulus presentations, may facilitate the emergence of this pattern of responding. That is, it may not be the fixed order of stimulus presentation per se, but the increased exposure to the test stimuli that generate the transformation of discriminative functions. For instance, in Experiment 9D, three out of four participants responded in accordance with the 5-member arbitrary comparative network when the test stimuli were presented in a fixed stimulus order, but participants were exposed to transformation training and testing for a total of two times. Similarly, in Experiments 9A and 9B, although participants were not re-exposed to training and test phases, they were exposed to an increased number of presentations of each of the relational stimuli during transformation tests. In turn, for the participants in Chapter 5 that failed to display the transformation of "More-than" and "Less-than" discriminative functions, it may have been beneficial to expose them to either an increased number of presentations of the test stimuli, or re-expose them to additional training and test phases. Future studies should therefore seek to undertake such an investigation.

Another potential reason for the differences observed in the current study, and those noted in the Dougher et al. (2007) study, may be due to the relational training procedures. For example, and as previously mentioned in Chapter 5, Stewart and

McElwee (2009) noted that in Experiment 1 of the Dougher et al. study, participants were successful in displaying non-arbitrary, and not arbitrary comparative responding to a 3-member relational network (A<B<C). In a subsequent phase of Experiment 1 in the Dougher et al. (2007) study, participants were trained to press the spacebar at a steady rate to stimulus B, and testing for the transformation of discriminative functions revealed that participants responded "less" to stimulus A, and "more" to stimulus C, than to stimulus B. Therefore, it may have been possible that participants in the Dougher et al. study, demonstrated the transformation of functions to non-arbitrary stimuli, and not arbitrary stimuli. However, Experiments 9A-9D in the current study appear to have overcome this limitation, in that a number of participants demonstrated the transformation of discriminative functions to an arbitrary comparative 5-member relational network. In doing so, the current study replicates and extends Dougher et al.'s findings. Furthermore, in the Dougher et al. study, participants were exposed to a respondent training phase, in which stimulus B was paired with mild electric shock and changes in skin conductance were recorded as the dependent measure. Findings revealed that participants showed "lower" changes in skin conductance to stimulus A, and "higher" changes to stimulus C, than to stimulus B. Thus, it may have been possible that the changes in skin conductance noted for a number of participants were to the non-arbitrary, and not arbitrary properties of the stimuli. Thus, it remains to be seen whether such changes in skin conductance would be observed if the current procedures were employed. This is an important issue that warrants further empirical investigation.

In summary, the findings from Experiments 9A-9D provide some tentative evidence for the transformation of discriminative functions to a 5-member arbitrary comparative network. Furthermore, the current findings appear to suggest that additional exposures to training and test phases, or an increased number of stimulus presentations during testing, may facilitate the emergence of this pattern of responding. However, further empirical research is needed on this issue. The implications of these findings will now be discussed.

Implications of findings

The findings from Chapter 5 of the current study may have important implications for our understanding of maladaptive behaviours, such as anxiety and avoidance (e.g., Dougher et al., 2007; Dymond et al., 2007, 2011). For example, a number of studies have proposed that the transfer, or transformation of functions, has the potential to model the development and maintenance of clinically significant behaviours (e.g., Dougher et al., 2007; Dymond et al., 2007, 2011; Roche et al., 2000). In one such study, Dymond et al. (2011) examined inferred threat-avoidance and safety behaviours (i.e. non-avoidance) following the establishment of responding to two equivalence classes (avoidance: AV1-AV2-AV3 and non-avoidance: NV1-NV2-NV3). During avoidance learning, a function was trained to one of the members of the avoidance class (e.g., AV2), and another was trained to a member of the non-avoidance (safety) class (e.g., NV2). That is, AV2 was paired with the presentation of an aversive image and sound, and NV2 was paired with the presentation of a pleasant image. However, during this phase, participants could learn to press the spacebar to cancel the upcoming image. Testing then involved presenting other members of both the inferred threat (i.e., avoidance) and safety (i.e., non-avoidance) equivalence classes. Findings demonstrated a significantly higher number of avoidance to the learned and inferred threat cues to the avoidance equivalence class members, in comparison to the learned and inferred safety cues in the non-avoidance class. Thus, on the basis of these findings, Dymond et al. (2011) propose that *arbitrarily applicable relational responding* has the potential to provide an alternative interpretation of how individuals come to avoid objects or events that they have never directly encountered. In turn, the findings from Chapter 5 of the current thesis are somewhat similar in that participants responded "more" to the stimuli ranked higher in the relational network (D and E), and "less" to the stimuli lower in the network (A and B), following training on the middle-ranked stimulus C. That is, participants demonstrated the predicted patterns of performance, in the absence of explicit reinforcement. Thus, the current findings have the potential to provide an alternative explanation as to how individuals may display increased or decreased levels of fear, or anxiety-related behaviours (see Dymond et al., 2011). However, further research examining the emergence of this pattern of responding with a "More-than/Less-than" avoidance paradigm is warranted.

Some potential limitations and suggestions for future research

Although the current thesis explored the potential utility of a number of different training and testing protocols in facilitating the emergence of TI-like behaviour and derived comparative responding, a number of potential limitations were noted. For example, in Experiments 5A and 5B in Chapter 3, and all experiments in Chapters 4 and 5, a relatively small participant population sample was employed. A total of ten and eight participants took part in Experiments 5A and 5B, respectively. However, the purpose of both of these experiments were to undertake a first investigation of the potential utility of a variant of the simple-to-complex protocol in generating arbitrary comparative responding in adult humans, and hence, the reason for the low number of participants. Again, across all experiments in Chapters 4 and 5, a small participant sample was employed. However, similar to the purpose of Experiments 5A and 5B in Chapter 3, the aim of Chapter 4 was to examine the potential utility of the RCP in establishing arbitrarily applicable comparative responding in adult participants, while Chapter 5 sought to undertake a first examination of the transformation of "More-than" and "Less-than" discriminative functions in accordance with a 5-member arbitrary relational. However, future studies should seek to incorporate a larger participant sample.

A further potential limitation to the current thesis was that the effectiveness of repeated exposures to training and testing phases in generating the targeted performances across Chapters 2-5, was not compared against a control group. Although, the protocol was found to have a facilitative effect for a number of participants who initially failed to display the predicted patterns of performance, it was not possible to compare performances against a group, who did not receive additional training and test phases. Future studies should therefore seek to undertake such a comparison.

As mentioned previously, although repeated exposure to training and test phases was found to have a facilitative effect on arbitrary comparative performances for a number of participants, a number of participants failed to demonstrate the predicted patterns of performance. In addition, the implementation of a range of interventions (e.g., variant of the simple-to-complex protocol, constructed-response protocol, MET consisting of non-arbitrary training trials), failed to remediate these deficiencies, and thus, it is worth considering some potential interventions that may be incorporated in future studies when such weaknesses are observed. For example, it may be beneficial for future studies seeking to examine the emergence of this behaviour, to incorporate an intervention similar to that employed in Experiment 3, in Chapter 2. That is, if participants were exposed to non-reinforced probe trials consisting of a number of one- and two-node combinatorially entailed relations during arbitrary relational training, this may have a facilitative effect on the emergence of arbitrarily applicable comparative responding. In addition, it may be beneficial to further increase the number of non-arbitrary training trials that participants are exposed to throughout the experimental task. For example, if additional non-arbitrary training trials were interspersed throughout arbitrary relational training blocks, this may help to foster the development of arbitrary comparative responding. Furthermore, Vitale et al. (2008) reported the utility of automated feedback (e.g., "Correct" and "Wrong") in facilitating responding to a number of 3-term problem-solving tasks in adult participants. More specifically, the authors found that feedback combined with non-arbitrary training trials, led to the largest improvements in participants' ability to respond to some of the more difficult problems-solving tasks. Therefore, it may have been beneficial in the current study to expose participants who continued to display weak inferential performances, to a brief intervention in which feedback was provided on some test relations. Future studies should therefore seek to undertake such an investigation.

With respect to Chapter 4, a potential limitation to this chapter was that no attempts were made to compare the effectiveness of the RCP to the conditional discrimination protocol. For example, a previous study by Dymond and Whelan (2010) undertook such an investigation, and found the RCP to be more successful in establishing "Same" and "Opposite" relations, than the MTS protocol. Although the current study did not undertake such a comparison, findings for participants that met criterion at testing in Chapter 4, appear to show superior performances on test relations, in comparison to those in Chapter 3, who were exposed to the conditional discrimination protocol. Therefore, it may be beneficial for future studies to undertake such an investigation.

Advantages of the current account

As mentioned, the current thesis sought to undertake an investigation of the utility of a novel behaviour-analytic account of TI-like behaviour in adult participants, based on the principles of Relational frame theory (RFT; Hayes et al., 2001a). In addition, a secondary aim of the current experimental work was to examine the utility of a range of interventions in facilitating the emergence of this pattern of responding, in adult humans. The potential advantages associated with the current account will now be discussed.

Relational frame theory is a contemporary account of human language and cognition that aims to provide a parsimonious analysis of the emergence of complex patterns of human behaviour (i.e., derived stimulus relations; Hayes et al., 2001a; Torneke, 2010). Much of this analysis takes place in controlled laboratory settings, in which researchers are provided with the opportunity to examine the conditions that contribute to the emergence of this type of responding. In addition, arbitrary stimuli are employed as contextual cues and comparison stimuli throughout arbitrary relational training and testing phases, so that it is possible to attribute the observed patterns of performance to intra-experimental contingencies arranged by the experimenter, and not to participants' prior histories of responding to comparative "More-than/Less-than" relations. Thus, the emergence of untrained patterns of arbitrarily applicable comparative responding that have been observed throughout numerous studies, provide an account of behaviour that is comparable to the emergence of TI responding. However, it must be noted that there is an important difference between the methods in which TI is currently examined, and those proposed by the current model.

For example, with the current account, participants are trained and tested using conditional discriminations, in which one of two contextual cues (MORE-THAN/LESS-THAN) are presented alongside two comparison stimuli on a given trial. Following repeated exposure to training trials, one comparison stimulus becomes discriminative for reinforcement (S+) in the presence of a particular contextual cue (i.e., conditional stimulus), and the other becomes discriminative for non-reinforcement (S-). That is, the function of the discriminative stimulus changes in the presence of the conditional stimulus (i.e., contextual cue) presented. Furthermore, with the current account, responding is first established to non-arbitrary comparative relations in the presence of the particular cues, and this pattern of behaviour is often found to generalise, such that, participants may respond to the relation between objects and events that are not based on physical properties (e.g., Hayes et al., 2001a; Torneke, 2010). This type of responding is referred to as *arbitrarily applicable relational responding (*AARR). Indeed, according to RFT, a fundamental ability of human cognition is the ability to relate events and stimuli (O'Toole, Barnes-Holmes, Murphy, O'Connor, & Barnes-Holmes, 2009). In comparison, with the method in which TI is currently examined, participants are exposed to a number of simultaneous discriminations where one stimulus from a pair becomes discriminative for reinforcement (S+), and the other becomes discriminative for non-reinforcement (S-). Despite the differences in training and testing between the aforementioned models, both allow an examination of the emergence of untrained patterns of performances.

With respect to the current account, it is possible to make precise predictions of which stimulus will be selected during a particular learning or test trial, on the basis of the contextual cue presented. For example, if stimuli A and B (B>A) are presented in the presence of the contextual cue for MORE-THAN, correct selections of B are predicted, whereas correct selections of A are predicted in the presence of the LESS-THAN contextual cue. This is in contrast to both reinforcement, and associative learning accounts (e.g., VTT) of TI, which predict that stimuli gain their value through a history of reinforcement and partial transfer of value from one stimulus to another during training (e.g., Frank et al., 2003; von Fersen et al., 1991; Wynne, 1995, 1998). Although RFT also states that a rich history of explicit reinforcement across numerous exemplars is central to the emergence of non-arbitrary and arbitrarily applicable relational responding, more precise control and accuracy in predictions, may be observed with the current contextualistic approach. In addition, the current predictions are also in contrast to those of cognitive accounts of TI. For instance, the Image theory proposes that the stimuli become integrated into a linear, hierarchical representation during training, which can then be accessed at testing, to make inferential judgments (de Soto et al., 1965). The current account however, does not make such predictions, and seeks to examine directly observable behaviour, without reference to unobservable structures.

A further potential advantage of the current account is that it is possible to examine the emergence of a wide range of test trials simultaneously. For example, current research studies investigating TI, examine the emergence of responding to non-endpoint and endpoint stimuli, alongside the trained relations (e.g., Acuna et al., 2002; Frank et al., 2005; Greene et al., 2001, 2006). This is also possible with the current account. However, as it is possible to train participants on a "More-than" relation and test them on a combination of "More-than" and "Less-than" relations, it is possible to examine an even greater number of performances at test with the current account. In turn, this is comparable to TI investigations of linear and non-linear trial types and thus, the current protocol may have the potential to further identify the problems that individuals encounter on some of the more difficult test trials (e.g., non-linear).

The current protocol also seeks to investigate the utility of a number of interventions that may be employed if weaknesses in relational performances are observed. This is further highlighted by the incorporation of mastery criterion during test phases, which allows the identification of deficiencies, or weaknesses in individuals' relational repertoires. Thus, with the current protocol, it is possible to deliver such interventions within the same test session, by means of a computer-generated program, which bodes a number of advantages. For example, if such relational training procedures were to be employed in the classroom setting, then it is possible for the program to re-cycle participants through training and test phases, without the need for teacher assistance (e.g., Connors et al., 1986). Furthermore, it is possible to tailor these programs to suit the needs of individuals.

Theoretical considerations

As mentioned in the previous section, the derived comparative account of TI proposed in the current thesis differs from those proposed by associative learning, and cognitive theorists, examining the TI problem. More specifically, all three theories differ in terms of the strategies individuals employ to solve test problems. For example, VTT proposes that differential stimulus values accrued during training, account for correct selections at test (von Fersen et al., 1991), whereas the Image theory proposes that mental representations of the stimuli allow individuals to solve problems at test (e.g., deSoto et al., 1965). In contrast, RFT proposes that verbal behaviour and the ability to respond in accordance with bidirectional, and combinatorially entailed relations, is critical to the emergence of TI.

A point worth considering regarding the proposed theories of TI centres on the linguistic capabilities of the populations studies. For example, the majority of studies examining TI involve adult humans with sophisticated verbal behaviour. With respect to associative learning accounts of TI, researchers propose that adult humans solve the task by employing similar strategies as non-humans (e.g., Frank et al., 2005; Greene et al., 2006; Wynne, 1997). That is, VTT proposes that adult humans solve tasks on the basis of a transfer of value from the reinforced stimulus (S+) to the unreinforced stimulus (S-). Although studies examining TI in adult humans employ non-verbal versions of the task, participants are typically exposed to detailed instructions at the start of the experiment. Thus, if humans are exposed to verbal instructions, it is questionable as to whether human performances on the TI task are comparable to non-humans. Indeed, proponents of VTT are very often silent on the fact that adult humans are exposed to detailed verbal instructions at the start of the experiment, which in turn, may influence responding during the task. For example, participants in the current thesis may have engaged in covert verbal behaviour throughout the course of experimental tasks. Thus, despite not being explicitly instructed to do so, participants may have assigned names or numbers to stimuli, in order to rank them in terms of their respective position in the relational network. So, for instance, participants may have assigned numbers to the stimuli, such as 1, 2, 3, 4 and 5. Therefore, on a particular learning trial, participants may have covertly said "2 beats 1" and/or "5 beats 1". At testing, participants may have relied on this same covert behaviour to solve novel test trials. If participants in the current thesis did employ covert behaviour during the task, then this may have influenced performances at test, which in turn, questions proposals that non-humans and humans solve TI problems using similar strategies.

Furthermore, a large proportion of studies examining TI in humans employ adults with sophisticated verbal behaviour and young children without learning, or language impairments. This in turn raises the issue as to how associative learning and cognitive theories of TI would deal with populations involving young children that have recognised language difficulties. For example, young children are capable of displaying TI when semi-verbal versions of the task are employed (e.g., Bryant & Trabasso, 1971). In a semi-verbal of the task, real objects are employed as stimuli, and language is used to describe the relationship between them (Wynne, 1998). However, young children with a recognised diagnosis of, for example, autism, may encounter difficulties responding to TI problems if language is used to describe the relationship between stimuli. Furthermore, as associative learning and cognitive theories of TI propose that associative values and mental representations established during training account for TI selections at test, then it is questionable as to how both theories would account for potential failures of individuals with language impairments to demonstrate TI. For example, Gorham et al. (2009) found that young children with a diagnosis of autism initially failed baseline tests, during which, probes for arbitrary comparative responding were presented. However, the authors reported that children with autism achieved criterion performances on these tasks, following the implementation of a multiple-exemplar training intervention, alongside experimenter-delivered verbal instructions. Thus, an important issue arising from these findings is that an examination of TI in populations that lack sophisticated verbal repertoires is warranted. That is, if individuals with language impairments failed to respond to TI problems, then such findings would have important implications for theoretical accounts of TI and the specific role of verbal behaviour on the emergence of this behaviour. Future studies should seek to undertake such investigations.

Conclusions

To conclude, the current thesis sought to determine the effectiveness of a number of variables in generating TI-like behaviour in adult participants. Findings from Chapter 2 suggest that awareness of the stimulus hierarchy leads to more accurate performances on the TI task. In addition, Chapter 2 highlighted the potential utility of adopting test mastery criterion and repeated exposures to training and testing phases, in examinations of TI. The effectiveness of a novel account of TI-like behaviour was examined throughout Chapters 3 and 4, and findings revealed that, for the most part, the protocol has the potential to generate arbitrary comparative responding, in adult participants. However, there are a number of issues outstanding with respect to the current training and testing protocols that warrant further investigation. For example, weaknesses were still noted for a number of participants in their relational repertoires following the implementation of a number of interventions. However, some of these limitations were overcome with the introduction of a novel procedure in Chapter 4, but further research is needed to explore the variables most conducive to the emergence of this pattern of responding. In addition, the current thesis sought to examine the transformation of discriminative functions to a 5-member arbitrary comparative network in Chapter 5. Findings from this chapter revealed that participants varied in their ability to respond in accordance with the relational network, and therefore, further empirical research is warranted on the conditions establishing this pattern of responding.

The protocols employed throughout Chapters 2-5 in the current thesis hold the potential to examine the emergence of TI in young children and individuals that lack sophisticated verbal repertoires. Indeed, the RCP protocol employed throughout Chapter 4 may hold the potential to be employed as an assessment tool examining the pre-requisites necessary for the demonstration of arbitrary comparative responding in young children, and individuals that lack sophisticated verbal repertoires. In addition, the current protocols have the potential to be incorporated in future studies seeking to examine the brain regions recruited during such tasks, and the cognitive impairments noted in clinical disorders, such as schizophrenia. Future studies should seek to explore these issues.

References

Acuna, B. D., Sanes, J. N., & Donoghue, J. P. (2002). Cognitive mechanisms of transitive inference. *Experimental Brain Research, 146,* 1-10.

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

Adkins, T., & Axelrod, S. (2001). Topography- versus selection-based responding: Comparison of mand acquisitions in each modality. *The Behaviour Analyst Today, 2,* 259-266.

Armstrong, K., Kose, S., Williams, L., Woolard, A., & Heckers, S. (2010). Impaired associative inference in patients with schizophrenia. *Schizophrenia Bulletin, 37,* 1–8.

Armstrong, K., Williams, L. E., & Heckers, S. (2012). Revised associative inference paradigm confirms relational memory impairment. *Neuropsychology, 26(4),* 451-458.

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

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

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.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes? *The Psychological Record, 50,* 251-265.

Barnes-Holmes, Y., Barnes-Holmes, D., & Cullinan, V. (2001). Education. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 181-196). New York: Kluwer Academic/Plenum.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry. *The Psychological Record, 51,* 287-308.

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Exemplar training and a derived transformation of function in accordance with symmetry: II. *The Psychological Record, 51*, 589-604.

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.

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for a relational operant. *Journal of Applied Behavior Analysis, 40,* 45-71.

Bernard, J., & Giurfa, M. (2004). A test of transitive inference in free-flying honeybees: Unsuccessful performance due to memory constraints. *Learning and Memory, 11,* 328-336.

Bond, A. B., Kamil, A. C., & Baldam, R. P. (2003). Social complexity and transitive inference in corvids. *Animal Behaviour, 65,* 479-487.

Bond, A. B., Wei, C. A., & Kamil, A. C. (2010). Cognitive representations in transitive inference: A comparison of four corvid species. *Behavioural Processes, 85(3),* 283-292.

Braine, M. D. S. (1998). Steps towards a mental-predicate logic. In M. D. S. Braine (Ed.), *Mental Logic* (pp. 273-332). New Jersey: Lawrence Erlbaum.

Breslow, L. (1981). Re-evaluation of the literature on the development of transitive inferences. *Psychological Bulletin, 89,* 325-351.

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature, 232,* 456-458.

Bush, R. R., & Mosteller, F. (1955). *Stochastic Models for Learning.* New York: Wiley.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564-575.

Cabello, F., Luciano, C., Gomez, I., & Barnes-Holmes, D. (2004). Human schedule performance, protocol analysis, and the "silent dog" methodology. *The Psychological Record, 54,* 405-422.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76,* 387-404.

Coleman, M. J., Titone, D., Krastoshevsky, O., Krause, V., Haans, Z., Mendell, N. R., Eichenbaum, H., & Levy, D. L. (2010). Reinforcement ambiguity and novelty do not account for transitive inference deficits. *Schizophrenia Bulletin, 36(6),* 1187-1200.

Colombo, M., & Frost, N. (2001). Representation of serial order in humans: A comparison to the findings with monkeys (Cebus apella). *Psychonomic Bulletin & Review, 8,* 262–269.

Connors, F. A., Caruso, D., & Detterman, D. (1986). *International Review of Research in Mental Retardation* (Vol. 14). New York: Academic Press.

Couvillin, P. A., & Bitterman, M. E. (1992). A conventional conditioning analysis of "transitive inference" in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 308-310.

Davis, H. (1992). Transitive inference in rats (Rattus norvegicus). *Journal of Comparative Psychology, 106,* 342-349.

Dawson, M. E., & Reardon, P. (1973). Construct validity of recall and recognition postconditioning measures of awareness. *Journal of Experimental Psychology, 98,* 308-315.

De Soto, C. B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology, 2,* 513-521.

De Souza, S., Goyos, C., Silvares, E. F. M., & Saunders, R. R. (2007). Emergence of printing and spelling skills from constructed-response matching-to-sample instruction. *European Journal of Behavior Analysis, 8,* 49-64.

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

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the Experimental Analysis of Behavior, 88,* 179-197.

Dougher, M. J., & Markham, M. R. (1994). Stimulus equivalence, functional equivalence, and the transfer of function. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition* (pp.71-91). Reno, NV: Context Press.

Dougher, M. J., & Markham, M. R. (1996). Stimulus classes and the untrained acquisition of stimulus function. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 71–90). Amsterdam: North Holland.

Dougher, M., Perkins, D. R., Greenway, D., Koons, A., & Chiasson, C. (2002). Contextual control of equivalence-based transformation of functions. *Journal of the Experimental Analysis of Behavior, 78,* 63-93.

Dube, W. V., McDonald, S. J., McIlvane, W. J., & Mackay, H. A. (1991). Constructed-response matching to sample and spelling instruction. *Journal of Applied Behavior Analysis, 24,* 305-317.

Dube W. V., & McIlvane W. J. (1999). Reduction of stimulus selectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis, 32,* 25–34.

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 75,* 5-22.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. *Proceedings of the National Academy of Sciences, 94,* 7109-7114.

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

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

Dymond, S., May, R. J., Munnelly, A., & Hoon, A. E. (2010). Evaluating the evidence base for relational frame theory: A citation analysis. *The Behavior Analyst, 33*, 97-117.

Dymond, S., Ng, T. C., & Whelan, R. (2013). Establishing arbitrarily applicable relations of same and opposite with the relational completion procedure: Selection-based feedback. *The Psychological Record, 63,* 1-20.

Dymond, S., & Rehfeldt, R. (2000). Understanding complex behaviour: The transformation of stimulus functions. *The Behavior Analyst, 23,* 239-254.

Dymond, S., Roche, B, & Barnes-Holmes, D. (2003). The continuity strategy, human behavior, and behavior analysis. *The Psychological Record, 53*, 333-347.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2007). Transformation of avoidance functions in accordance with the relational frames of same and opposite. *Journal of the Experimental Analysis of Behavior, 88,* 249-262.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2008). Derived avoidance learning: Transformation of avoidance response functions in accordance with the relational frames of same and opposite. *The Psychological Record, 58,* 271-288.

Dymond, S., Schlund, M., Roche, B., Whelan, R., Richards, J., & Davies, C. (2011). Inferred threat and safety: Symbolic generalization of human avoidance learning. *Behaviour Research and Therapy, 49*, 614-621.

Dymond, S., & Whelan, R. (2010). Derived relational responding: A comparison of match-to-sample and the relational completion procedure. *Journal of the Experimental Analysis of Behavior, 94,* 37-55.

Dymond, S., Whelan, R., & Smeets, P. M. (2005). A transformation of discriminative functions in accordance with equivalence relations. *European Journal of Behaviour Analysis, 6*, 111-123.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences of the United States of America, 18,* 7723-7728.

Escobar, R., & Bruner, C. A. (2008). Effects of the contiguity between the extinction and the reinforcement components in observing-response procedures. *Mexican Journal of Behavior Analysis, 34,* 333–347.

Escobar, R., & Bruner, C. A. (2009). Observing responses and serial stimuli: Searching for the reinforcing properties of the S-. *Journal of the Experimental Analysis of Behavior, 92,* 215-231.

Evans, T., Newstead, S. E., & Byrne, R. M. (1993). *The Psychology of Deduction.* Hillsdale, England: Lawrence Erlbaum.

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

Fields, L., Varelas, A., Reeve, K. F., Belanich, J., Wadhwa, P., de Rosse, P., & Rosen, D. (2000). Effects of prior conditional discrimination training, symmetry, transitivity, and equivalence testing on the emergence of new equivalence classes. *The Psychological Record, 50,* 443-466.

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

Fienup, D. M., Covey, D. P., & Critchfield, T. S. (2010). Teaching brain-behavior relations economically with stimulus equivalence technology. *Journal of Applied Behavior Analysis, 43,* 19-33.

Fletcher, J. D. (1990). *Effectiveness and cost of inter-active videodisc instruction in defense training and education* (IDA Paper No. P-2372). Alexandria, VA: Institute for Defense Analyses.

Frank, M. J., Rudy, J. W., Levy, W. B., & O'Reilly, R. C. (2005). When logic fails: Implicit transitive inference in humans. *Memory and Cognition, 33,* 742-750.

Frank, M. J., Rudy, J. W., & O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. *Hippocampus, 13,* 341-354.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science, 306,* 1940–1943.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory, 13,* 259–262.

Gazes, R. P., Chee, N. W., Hampton, R. R. (2012). Cognitive mechanisms for transitive inference performance in rhesus monkeys: Measuring the influence of associative strength and inferred order. *Journal of Experimental Psychology: Animal Behavior Processes, 38(4),* 331-345.

Gigerenzer, G. (1997). Bounded rationality: Models of fast and frugal inference. *Swiss Journal of Economics and Statistics, 133(2/2),* 201-218.

Gil, E., Luciano, C., Ruiz, F. J., & Valdivia-Salas, V. (2012). A preliminary demonstration of the transformation of functions through hierarchical relations. *International Journal of Psychology and Psychological Therapy, 12(1),* 1-19.

Gillan, D. J. (1981). Reasoning in the chimpanzee: II. Transitive inference. *Journal of Experiment Psychology: Animal Behavior Processes, 7,* 150-164.

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.

Greene, A. J., Gross, W. L., Elsinger, C. L., & Rao, S. M. (2006). An fMRI analysis of the human hippocampus: Inference, context, and task awareness. *Journal of Cognitive Neuroscience, 18,* 1156-1173.

Greene, A. J., Spellman, B. A., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: Transitive inference using nonverbal stimuli in humans. *Memory and Cognition, 29,* 893-902.

Gross, A. C., & Fox, E. J. (2009). Relational frame theory: An overview of the controversy. *Analysis of Verbal Behavior, 25(1),* 87-98.

Hayes, S. C. (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control.* New York: Plenum.

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

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior, 68,* 235-244.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/Plenum Publishers.

Hayes, S. C., Brownstein, A. J., Haas, J. R., & Greenway, D. E. (1986). Instructions, multiple schedules, and extinction: Distinguishing rule governed from schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 46,* 137-147.

Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2001). Derived relational responding as learned behaviour. In S. C. Hayes, D. Barnes-Holmes, & Roche, B. (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 21-49). New York: Kluwer Academic/Plenum Publishers.

Heckers, S., Zalesak, M., Weiss, A. P., Ditman, T., & Titone, D. (2004). Hippocampal activation during transitive inference in humans. *Hippocampus, 14,* 153-162.

Hinton, E. C., Dymond, S., Von Hecker, U. & Evans, C. J. (2010). Neural correlates of relational reasoning and the symbolic distance effect: Involvement of parietal cortex. *Neuroscience, 168,* 138-148.

Hunter, I. M. L. (1957). The solving of the three term series problems. *British Journal of Psychology, 48,* 286-298.

Huttenlocher, J. (1968). Constructing images: A strategy in reasoning. *Psychological Review, 75,* 550-560.

Hyland, J., O'Hora, D., Smyth, S., & Leslie, J. C. (2012). Sequential responding in accordance with temporal relational cues: A comparison of "Before" and "After". *The Psychological Record, 62,* 463-484.

Johnson-Laird, P. N. (1972). The three-term series problem. *Cognition, 1(1),* 57-82.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Erlbaum.

Koscik, T. R., & Tranel, D. (2012). The human ventromedial prefrontal cortex is critical for transitive inference. *Journal of Cognitive Neuroscience, 5(1),* 191-204.

Kritch, K. M., & Bostow, D. E. (1994). Creating computer programmed instruction [Computer program]. Tampa, FL: Customs Systems International, Inc.

Lazareva, O. F., Smirnova, A. A., Bagozkaja, M. S., Zorina, Z. A., Ravevsky, V. V., & Wasserman, E. A. (2004). Transitive responding in hooded crows requires linearly ordered stimuli. *Journal of the Experimental Analysis of Behavior, 82,* 1-19.

Lazareva, O. F., Smirnova, A. A., Zorina, Z. A., & Rayevsky, V. V. (2001). Hooded crows solve a transitive inference problem cognitively. *Animal Welfare, 10,* 219-231.

Lazareva, O. F., & Wasserman, E. A. (2010). Nonverbal transitive inference: Effects of task and awareness on performance. *Behavioural Processes, 83,* 99-112.

Leo, P. D., & Greene, A. J. (2008). Is awareness necessary for true inference? *Memory & Cognition, 36,* 1079-1086.

Libben, M., & Titone, D. (2008). The role of awareness and working memory in human transitive inference. *Behavioural Processes, 77,* 43-54.

Lieving, G. A., Reilly, M. P., & Lattal, K. A. (2006). Disruption of responding maintained by conditioned reinforcement: alterations in response–conditioned reinforcer relations. *Journal of the Experimental Analysis of Behavior, 86,* 197–209.

Lovibond, P. F., & Shanks, P. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes, 28(1),* 3-26.

Lowenkron, B. (1991). Joint control and the generalization of selection-based verbal behavior. *Analysis of Verbal Behavior, 9,* 121-126.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Luciano, M. C., Gomez-Becerra, I. G., & Valverde, M. R. (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.

Mackay, H. A., & Fields, L. (2009). Syntax, grammatical transformation and productivity: A synthesis of stimulus sequences, equivalence classes and contextual control. In R. A. Rehfeldt, & Y. Barnes-Holmes (Eds.), *Derived relational responding: Applications for learners with autism and other developmental disabilities* (pp. 209-235). Oakland, CA: New Harbinger Publications.

Manns, J. R., & Squire, L. R. (2001). Perceptual learning, awareness and the hippocampus. *Hippocampus, 11(6),* 776-782.

Markham, M. R., & Dougher M. J. (1994). Compound stimuli in emergent stimulus relations: Extending the scope of stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 60,* 529-542.

Martin, N., & Alsop, B. (2004). Transitive inference and awareness in humans. *Behavioural Processes, 67,* 157-165.

McGonigle, B., & Chalmers, M. (1977). Are monkeys logical? *Nature,* 267, 694-696.

McGonigle, B. O., & Chalmers, M. (1984). The selective impact of question form and input code on the symbolic distance effect in children. *Journal of Experimental Child Psychology, 37,* 525-554.

Merritt, D. J., & Terrace, H. S. (2011). Mechanisms of inferential order judgments in humans (Homo Sapiens) and rhesus monkeys (Macca Mulatta). *Journal of Comparative Psychology, 125(2)*, 222-238.

Michael, J. (1985). Two kinds of verbal behavior and a possible third. *The Analysis of Verbal Behavior, 3,* 2-5.

Michael, J. (1993). Establishing operations. *The Behavior Analyst, 16,* 191-206.

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2010). Brief report: The training and assessment of relational precursors and abilities (TARPA): A preliminary analysis. *Journal of Autism and Developmental Disorders, 40,* 1149-1153.

Moses, S. N., Villate, C., Binns, M. A., Davidson, P. S. R., & Ryan, J. D. (2008). Cognitive integrity predicts transitive inference performance bias and success. *Neuropsychologia, 46,* 1314–1325.

Moses, S. N., Villate, C., & Ryan, J. D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. *Neuropsychologia, 44,* 1370-1387.

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8,* 228-246.

Munnelly, A., Dymond, S., & Hinton, E. C. (2010). Relational reasoning with derived comparative relations: A novel model of transitive inference. *Behavioural Processes, 85,* 8-17.

Nagode, J. C., & Pardo, J. V. (2002). Human hippocampal activation during transitive inference. *Neuroreport, 13,* 939-944.

O'Hora, D., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2004). Derived relational networks and control by novel instructions: A possible model of generative verbal responding. *The Psychological Record, 54,* 437-460.

O'Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *The Psychological Record, 52,* 51-75.

O'Toole, C., Barnes-Holmes, D., Murphy, C., O'Connor, J., & Barnes-Holmes, Y. (2009). Relational flexibility and human intelligence: Extending the remit of Skinner's verbal behavior. *International Journal of Psychology and Psychological Therapy, 9,* 1-17.

Piaget, J. (1928). *Judgment and reasoning in the child.* Harcourt, Brace and Co: New York.

Potter, B., & Brown, D. L. (1997). Review of the studies examining the nature of selection-based and topography-based verbal behavior. *The Analysis of Verbal Behavior, 2,* 85-104.

Potts, G. R. (1974). Storing and retrieving information about ordered relationships. *Journal of Experimental Psychology, 103,* 431-439.

Rapp, P. R., Kansky, M. T., & Eichenbaum, H. (1996). Learning and memory for hierarchical relationships in the monkey: Effects of aging. *Behavioral Neuroscience, 110,* 887-897.

Reilly, T., Whelan, R., & Barnes-Holmes, D. (2005). The effect of training structure on the latency of responses to a five-term linear chain. *The Psychological Record, 55,* 233-249.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokosy (Eds.), *Classical Conditioning II: Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Riley, C., & Trabasso, T. (1974). Comparatives, logical structures and encoding in a transitive inference task. *Journal of Experimental Child Psychology, 17,* 187-203.

Roberts, W. A., & Phelps, M. T. (1994). Transitive inference in rats: A test of the spatial coding hypothesis. *Psychological Science, 5,* 368–374.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus functions in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior, 67,* 275–301.

Roche, B., Barnes-Holmes, D., Smeets, P. M., Barnes- Holmes, Y., & McGeady, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record, 50,* 267–291.

Rosenfarb, I. S., Newland, M. C., Brannon, S. E., & Howey, D. S. (1992). Effects of self-generated rules on the development of schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 58,* 107-121.

Russell, J., McCormack, T., Robinson, J., & Lillis, S. G. (1996). Logical (versus associative) performance on transitive reasoning tasks by children: Implications for the status of animals' performance. *Quarterly Journal of Experimental Psychology, 49B,* 231-244.

Ryan, J. D., Moses, S. N., & Villate, C. (2009). Impaired relational organization of propositions but intact transitive inference in aging: Implications for understanding underlying neural integrity. *Neuropsychologia, 47,* 338-353.

Saunders, K., & Williams, D. (1998). Stimulus control procedures. In K. A. Lattal, & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 193-228). New York: Plenum.

Schacter, D. L. (1998). Memory and awareness. *Science, 280,* 59- 60.

Schaffer, L., & Hannafin, M. J. (1986). The effects of progressively enriched interaction on learning from interactive video. *Educational Communication and Technology Journal, 34,* 89–96.

Sedek, G., & von Hecker, U. (2004). Effects of subclinical depression and aging on generative reasoning about linear orders: Same or different processing limitations? *Journal of Experimental Psychology: General, 133,* 237-260.

Shafer, E. (1993). Teaching topography-based and selection-based verbal behavior to developmentally disabled individuals: Some considerations. *The Analysis of Verbal Behavior, 11,* 117-133.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5-13.

Sidman M. (1994). *Equivalence relations and behavior: A research story.* Boston, MA: Authors Cooperative.

Siemann, M., & Delius, J. D. (1993). Implicit deductive responding in humans. *Naturwissenschaften, 80,* 364-366.

Siemann, M., & Delius, J. D. (1996). Influences of task concreteness upon transitive inference in humans. *Psychological Research, 59,* 81-93.

Siemann, M., & Delius, J. D. (1998). Algebraic learning and neural network models for transitive and non-transitive responding. *European Journal of Cognitive Psychology 10,* 307–334.

Simon, H. A. (1957). *Models of man, social and rational: Mathematical essays on rational human behavior in a social setting.* New York: Wiley.

Simon, H. A. (1996a). *Models of my life. MIT Edition.* Cambridge, MA: MIT Press.

Simon, H. A. (1996b). *The sciences of the artificial.* Cambrdige, MA: MIT Press.

Smeets, P. M., Barnes-Holmes, D., & Striefel, S. (2006). Establishing and reversing equivalence relations with precursor to the relational evaluation procedure. *The Psychological Record, 56(2),* 267-286.

Smeets, P. M., van Wijngaarden, M., Barnes-Holmes, D., & Cullinan, C. (2004). Assessing stimulus equivalence with a precursor to the relational evaluation procedure. *Behavioral Processes, 65,* 241-251.

Smith, C., & Squire, L. R. (2005). Declarative memory, awareness, and transitive inference. *Journal of Neuroscience, 25,* 10138-10146.

Steirn, J. N., Weaver, J. E., & Zentall, T. R. (1995). Transitive inference in pigeons: Simplified procedures and a test of value transfer theory. *Animal Learning & Behavior, 23,* 76-82.

Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the relational evaluation procedure. *The Psychological Record, 54*, 531-552.

Stewart, I., & McElwee, J. (2009). Relational responding and conditional discrimination procedures: An apparent inconsistency and clarification. *The Behavior Analyst, 32*, 309-317.

Sundberg, M. L., & Michael, J. (2001). The value of Skinner's analysis of verbal behavior for teaching children with autism. *Behavior Modification*, *25,* 698-724.

Sundberg, C. T., & Sundberg, M. L. (1990). Comparing topography-based verbal behavior with stimulus selection-based verbal behavior. *Analysis of Verbal Behavior, 8,* 31-41.

Titone, D., Ditman, T., Holzman, P. S., Eichenbaum, H., & Levy, D. L. (2004). Transitive inference in schizophrenia: Implications in relational memory organization. *Schizophrenia Research, 68,* 235-247.

Torneke, N. (2010). *Learning RFT: An introduction to relational frame theory and its clinical applications.* Oakland, CA: New Harbinger Publications, Inc.

Trabasso, T. (1975). Representation, memory, and reasoning: How do we make transitive inferences? In A. D. Pick (Ed.), *Minnesota symposia on child psychology* (pp. 135-172). Minneapolis University: Minnesota Press.

Trabasso, T. (1977). The role of memory as a system in making transitive inference. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 333-366). Hillsdale, NJ: Erlbaum.

Trabasso, T., & Riley, C. A. (1975). The construction and use of representations involving linear order. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 381-410). Hillsdale, NJ: Erlbaum.

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), *Reasoning: Representation and processes* (pp. 201-229). Hillsdale, NJ: Erlbaum.

Treichler, F., & vanTilberg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macque monkeys. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 105-117.

Van Elzakker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus. I. An empirical analysis. *Hippocampus, 13,* 334-340.

Van Opstal, F., Verguts, T., Orban, G. A., & Fias, W. (2008). A hippocampal-parietal network for learning an ordered sequence. *NeuroImage, 40,* 333-341.

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes, 78,* 313-334.

Vignes, T. (2007). A comparison of topography-based and selection-based verbal behaviour in typically developed children and developmentally disable persons with autism. *Analysis of Verbal Behavior, 23,* 113-122.

Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitating responding in accordance with the relational frame of comparison: Systematic empirical analyses. *The Psychological Record, 58,* 365-390.

Von Fersen, L., Wynne, C., Delius, J. D., & Staddon, J. (1991). Transitive inference in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 334-341.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In R. A. Boakes, & M. S. Haliday (Eds.), *Inhibition and learning* (pp.301-336). New York: Academic Press.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boon, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10,* 119–125.

Weaver, J. E., Steirn, J. N., & Zentall, T. R. (1997). Transitive inference in pigeons: Control for differential value transfer. *Psychonomic Bulletin & Review, 4,* 113-117.

Wendelken, C., & Bunge, S. A. (2010). Transitive inference: Distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *Journal of Cognitive Neuroscience, 22,* 837-847.

Whelan, R., & Barnes-Holmes, D. (2004). The transformation of consequential functions in accordance with the relational frames of Same and Opposite. *Journal of the Experimental Analysis of Behavior, 82,* 177-195.

Whelan, R., Barnes-Holmes, D., & Dymond, S. (2006). The transformation of consequential functions in accordance with the relational frames of more-than and less-than. *Journal of the Experimental Analysis of Behavior, 86,* 317-335.

Williams, L. E., Avery, S. N., Woolard, A. A., & Heckers, S. (2012). Intact relational memory and normal hippocampal structure in the early stage of psychosis. *Biological Psychiatry, 71,* 105-113.

Willingham, D. B. (1997). Implicit and explicit memory do not differ in flexibility: Comment on Dienes and Berry (1997). *Psychonomic Bulletin & Review, 4,* 587-591.

Wyckoff, L. B., Jr. (1952). The role of observing responses in discrimination learning. Part I. *Psychological Review, 59,* 431–442.

Wyckoff, L. B., Jr. (1969). The role of observing responses in discrimination learning. Part II. In D. P. Hendry (Ed.), *Conditioned reinforcement* (pp. 237–260). Homewood, IL: Dorsey Press.

Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning & Behavior, 23,* 207-217.

Wynne, C. D. L. (1997). Pigeon transitive inference: Tests of simple accounts of a complex performance. *Behavioural Processes, 19,* 95-112.

Wynne, C. D. L. (1998). A minimal model of transitive inference. In C.D.L. Wynne and J. E. R. Staddon (Eds.), *Models of action: Mechanisms for adaptive behavior* (pp.269-307). Mahwah, NJ: Erlbaum.

Wynne, C. D. L., von Fersen, L., & Staddon, J. E. R. (1992). Pigeons' inference are transitive and the outcome of elementary conditioning principles: A response. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 313-315.

Zalesak, M., & Heckers, S. (2009). The role of the hippocampus in transitive inference. *Psychiatry Research: Neuroimaging, 172,* 24-30.

Appendices

Appendix 1

Sample Participant Information Sheet

In this study, you will receive a series of instructions and tasks presented on a computer. You are asked to read the instructions and to follow them to the best of your ability.

Tasks will merely involve you making choices between objects presented on-screen. You will make choices by using the computer keyboard. Sometimes you will receive feedback on your choices and sometimes you will not. However, it is possible to get the tasks without feedback correct by paying careful attention to the feedback you receive during the tasks with feedback.

You will be provided with a minimum of 3 of subject pool credits on completion of the study. If the task takes longer, you will receive the appropriate number of credits.

I will provide you with a full debriefing, and answer any questions that you might have, at the end of the study. Your rights as a participant, including the right to withdraw at any point without penalty, are ensured.

If you have any questions at all, please ask them now. If you would like to participate, please ask the researcher for a consent form.

Please contact for further information: Anita Munnelly (492834@swansea.ac.uk)

Appendix 2

Sample Consent Form

Name of Participant:________________Date: ___________ Age: __________

I consent that I am willing to participate in this study.

I am satisfied with the instructions I have been given so far and I expect to have any further information requested regarding the study supplied to me at the end of the experiment. I will not interact with the experimenter during the experiment unless I wish to terminate my participation.

I have not been coerced in any way to participate in this study and I understand that I may terminate my participation in the study at any point should I so wish. I am over 18 years of age.

Name of participant

(print).................................Signed...................Date............

Name of researcher

(print)................................Signed...................Date...........

Appendix 3

Sample Debriefing Form

This study was designed to examine transitive inference (e.g. problem-solving). You were initially trained with a series of learning tasks using nonsense words in which feedback was delivered every time you made a correct response. Following that, you were tested without any feedback or reinforcement in order to measure the degree to which you had learnt the relations. We are interested in the accuracy of people's learning during these tasks, which is called 'transitive inference'. For instance, if you learn that A>B and B>C, then you may also learn, without feedback, that A>C.

I hope that this has helped to clarify for you the purpose of the study you have just undertaken.

Your participation in the study is greatly appreciated; thank you!

Appendix 4

Post-experimental questionnaire

1. What did you think we were trying to find out in this experiment?

2. In your opinion, were all of the pairs in the no-feedback condition the same as the pairs in the condition where you were given feedback (Please circle one of the following)?

Yes No Not Sure

If no, do you think there was a correct answer?

Yes No Not Sure

3. If you believe there was a correct answer, explain why:

a. There is a logically correct answer because (explain):

b. One just seemed right (explain why):

c. I guessed there may be a correct answer but I don't know what it is.

d. Other (explain):

4. You were presented with the following images several times, but never told the correct answer. Circle one of the images below that you think is correct (guess if necessary):

哲 安

Appendix 4 (*cont.d*)

5. You were presented with the following images several times, but never told the correct answer. Circle one of the images below that you think is correct (guess if necessary):

6. You were presented with the following images several times, but never told the correct answer. Circle one of the images below that you think is correct (guess if necessary):

7. For questions 4, 5, and 6 above, what reason (if any) did you use to learn the images (circle one):

a. I already know the images: If so, from where?

b. I gave them names.

c. I memorised part of each image.

d. I just watched and eventually got it.

e. I used their similarity to familiar shapes.

f. No strategy.

g. Other (explain):

Appendix 4 (*cont.d*)

8. Based on your understanding of how the images relate to one another, arrange the images appropriately below, using the numbers assigned to each.

1. 哲 **2.** 室 **3.** 安

4. 客 **5.** 震 **6.** 南

i.

ii.

iii.

iv.

v.

vi.

9. When did you become aware of the relationship between the images (circle one):

a. During the phases with feedback.

b. During the first few phases without feedback.

c. During the final phase without feedback (last section completed).

Appendix 5

List of the non-arbitrary stimulus sets employed throughout Experiments 6A-9D (Chapters 4 and 5).

Non-arbitrary Relational Training and Testing	Non-arbitrary Constructed-Response Relational Training and Testing
Apples (1, 2, 4, 5, 6, 8)	Pigs (1, 2, 3, 4, 5, 6)
Basketballs (1, 2, 3, 4, 6, 8)	Guitars (1, 2, 3, 4, 5, 6)
Books (1, 4, 5, 6, 7, 8)	Planes (1, 3, 4, 5, 7, 9)
Cars (1, 2, 3, 4, 5, 6)	Sheep (2, 4, 5, 6, 7, 8)
Butterflies (1, 4, 5, 6, 7, 8)	Snowmen (1, 2, 3, 4, 5, 6)
Chemical Flasks (1, 3, 4, 6, 7, 8)	Saw (3, 4, 5, 6, 7, 8)
Clocks (1, 2, 3, 4, 5, 6)	Cowboys (1, 2, 3, 4, 5, 6)
Cubes (1, 2, 3, 4, 5, 6)	Birds (1, 2, 3, 4, 5, 6)
Dogs (1, 2, 3, 4, 5, 6)	Computers (3, 4, 5, 6, 7, 8)
Doughnuts (1, 2, 3, 4, 5, 6)	Scissors (1, 3, 5, 6, 7, 9)
Fish (1, 4, 5, 6, 7, 8)	Chickens (3, 4, 5, 6, 7, 8)
Lights (1, 2, 3, 4, 5, 6)	Ducks (1, 2, 3, 4, 5, 6)
Arks (1, 2, 3, 4, 5, 6)	Cups (1, 2, 3, 5, 6, 8)
Phones (1, 2, 3, 4, 5, 6)	Bicycles (1, 2, 3, 4, 5, 6)
Ships (1, 2, 3, 4, 5, 6)	Fire-trucks (1, 3, 4, 6, 7, 8)
Squiggles (1, 2, 4, 6, 7, 8)	Sharks (1, 2, 3, 4, 5, 6)
Stars (1, 2, 3, 4, 5, 6)	Buckets (1, 2, 3, 5, 6, 7)
Tractors (1, 2, 3, 4, 5, 6)	Umbrellas (3, 4, 5, 6, 7, 8)
Trees (1, 2, 3, 4, 5, 6)	Tennis balls (2, 3, 4, 5, 6, 7)
Carrots (2, 3, 4, 5, 6, 7)	Baseball bats (3, 4, 5, 6, 7, 10)
Chairs (2, 3, 4, 5, 6, 7)	Pears (1, 2, 3, 4, 5, 6)
Giraffes (1, 2, 3, 4, 5, 6)	Racing Cars (1, 2, 3, 4, 5, 6)
Lady-birds (2, 4, 5, 6, 7, 8)	Violins (2, 3, 4, 5, 6, 7)
Prams (1, 2, 4, 5, 6, 8)	Radios (3, 4, 5, 6, 7, 8)
Sun (2, 3, 4, 6, 7, 8)	Wine bottles (2, 3, 4, 5, 6, 7)
Turtles (1, 2, 3, 4, 5, 6)	Squirrels (1, 2, 3, 4, 5, 6)
Bats (1, 2, 3, 4, 6, 8)	Spaceships (1, 2, 3, 4, 5, 6)
Circles (3, 4, 5, 6, 7,8)	Helicopters (3, 4, 5, 6, 7, 8)