Quantifying differential stimulus relatedness using the Function Acquisition Speed Test

Thesis submitted to the Department of Psychology, Faculty of Science, in fulfilment of the requirements for the degree of Master of Science, National University of Ireland, Maynooth.

Jamie Cummins, B.A. (Hons)

October 2017

Research supervisor: Dr. Bryan Roche

Head of Department: Prof. Andrew Coogan

Contents

Table of Contents……………………………………………………………………………..ii

List of Tables...vi

List of Figures..vii

List of Equations..viii

Declaration…………………………………………………………………..………..ix

Acknowledgements……………………………………...……………………......…...x

Abstract…………………………………………………………………..……………...xi

Chapter 1: Introduction…………………………………………………………...1

1.1 Introduction……………………………………………………………………….2

1.2 A Brief History of Stimulus Equivalence…………………………………………………2

1.2.1 Early Behaviour-Analytic Approaches to Language……………………………2

1.2.2 Sidman's Paradigm……………………………………………………………...3

1.2.3 Stimulus Equivalence and Language……………………………………………4

1.3 Stimulus Equivalence and Differential Response Strengths......……………………………5

1.3.1 Stimulus Equivalence is a Behavioural Process..………………………………5

1.3.2 Resistance to change……………………………………....…………………6

1.3.3 Stimulus Relatedness…………………………………………………………8

1.4 Quantifying Stimulus Relatedness……………………………………………………10

1.5 Background to the Function Acquisition Speed Test……………………………………13

1.5.1 Implicit Measures and the IAT…………………………………………………13

1.5.2 What Does the IAT Measure?...15

1.5.3 The Implicit Relational Assessment Procedure…………….……………………..17

1.5.4 Procedural and Conceptual issues with the IRAP……………………………..19

1.6 The Function Acquisition Speed Test…………………………..……………23

1.6.1 The FAST Procedure....………………………………………..……………24

1.6.2 Current FAST Scoring…………………………………..…………..25

1.6.3Variations of the FAST Format..26

1.7 The FAST and Stimulus Relatedness……………………………………………..28

1.7.1 Measuring Laboratory Controlled Stimulus Relations…………………29

1.7.2 The Current Research…………………………………………………..32

Chapter 2: Assessing the FAST as a Means of Detecting Differential Stimulus Relatedness along Inter-Class Parameters. Experiment 1……………………………35

2.1 Introduction……………………………………………………………………36

2.2 Method…………………………………………………………………………37

2.2.1 Subjects……………………………………………………………………37

2.2.2 Ethical Considerations…………………………………………………….37

2.2.3 Apparatus…………………………………………………………………38

2.2.4 Procedure…………………………………………………………………39

2.2.4.1 General Sequence…………………………………………………39

2.2.4.2 Stimulus Equivalence Training and Testing……………………40

2.2.4.3 Function Acquisition Speed Test………………………………42

2.3 Results…………………………………………………………………………44

2.3.1 Measurement………………………………………………………………44

2.3.2 Descriptive Statistics………………………………………………………46

2.3.2.1 Matching-to-Sample………………………………………………46

2.3.2.2 Function Acquisition Speed Test………………………………46

2.3.3 Inferential Statistics………………………………………………………49

2.3.3.1 Frequentist Analyses………………………………………………49

2.3.3.2 Bayesian Analyses…………………………………………………51

2.3.4 Summary……………………………………………………………………53

2.4 Discussion………………………………………………………………………53

Chapter 3: Assessing the FAST as a Means of Detecting Differential Stimulus Relatedness along Intra-Class Parameters. Experiment 2……………………………60

3.1 Introduction……………………………………………………………………61

3.2 Method……………………………………………………………………………...62

3.2.1 Subjects……………………………………………………………………62

3.2.2 Ethical Considerations…………………………………………………62

3.2.3 Apparatus………………………………………………………………63

3.2.4 Procedure………………………………………………………………63

3.2.4.1 General Sequence……………………………………………….63

3.2.4.2 Stimulus Equivalence Training.................……………………….64

3.2.4.3 Function Acquisition Speed Test………………………………...67

3.2.4.4 Stimulus Equivalence Testing...67

3.3 Results……………………………………………………………………………69

3.3.1 Measurement……………………………………………………………...69

3.3.2 Descriptive Statistics……………………………………………………...69

3.3.2.1 Stimulus Equivalence Training...69

3.3.2.2 Function Acquisition Speed Tests...70

3.3.2.3 Stimulus Equivalence Testing...73

3.3.3 Inferential Statistics………………………………………………………75

3.3.3.1 Frequentist Analyses……………………………………………...75

3.3.3.2 Bayesian Analyses…..…………………………………….……77

3.3.4 Summary...77

3.4 Discussion…………..……………………………………………………………77

Chapter 4: General Discussion..83

4.1 Research Summary and Main Findings...84

4.1.1 Quantifying Stimulus Relatedness...84

4.1.2 Conceptualising and Controlling Relatedness...87

4.1.3 Implicit Test Effect Sizes and Comparability of Procedures...........................89

4.2 The FAST and Single-Subject Measurement...91

4.2.1 Effects of Order...92

4.2.2 Effects of Practice..94

4.3 Quantifying Fluency in the FAST...96

4.4 Quantifying Emergent Stimulus Classes..99

4.5 Potential Confounds of the Current Research...101

4.5.1 Effect of Contrast Stimuli...101

4.5.2 Limits of MTS in Controlling Stimulus Relations..104

4.6 Expanding the Continuum of Scores in the FAST..106

4.7 Summary and Conclusion..109

References...112

Appendices..131

List of Tables

Table 1: Comparison of the IAT and IRAP..20

Table 2: Comparison of the IAT, IRAP, and FAST...27

Table 3: Mean differences in scores between blocks for subjects in Experiment 1.................46

Table 4: Mean scores for subjects in both the consistent and consistent FAST block in Experiment 1..48

Table 5: Data for each subject in the MTS training phase of Experiment 2............................70

Table 6: FAST and RFD scores for each subject in each FAST of Experiment 2.................72

Table 7: Accuracy and response time difference scores for each subject in each FAST of Experiment 2..72

Table 8: Data for each subject in the MTS testing phase of Experiment 2..............................74

Table 9: Comparison of FAST score, RFD score, and MTS testing performances of each subject in Experiment 2..74

List of Figures

Figure 1: Explanation of intra-class and inter-class stimulus relations....................................10

Figure 2: Layout of the IAT procedure...14

Figure 3: Schematic of FAST Response Format..44

Figure 4: Changes in mean (a) FAST score, and (b) RFD score across conditions in Experiment 1..47

Figure 5: Changes in the consistent and inconsistent block of (a) FAST score, and (b) RFD across conditions in Experiment 1...51

Figure 6: Changes in the scores of each subject in Experiment 2 on each FAST...................71

List of Equations

Equation 1: Rate-fluency differential formula..45

Equation 2: FAST block-slope formula...56

Declaration

I, the undersigned, hereby certify that this material, which I now submit in fulfilment of a M.Sc. degree, has not been previously submitted as an exercise for a degree at this or any other University, and is, unless otherwise stated, entirely my own work.

Signed:________________________________

Jamie Cummins

13314401

Date:

Acknowledgements

Firstly, to my supervisor, mentor, and friend, Dr. Bryan Roche: it's impossible to overstate how influential you've been in my formative years in academia. Thank you for all of the opportunities you've provided me, the "20 minute drop-ins" that started at 2pm and ended 4 hours later, and for letting me learn from your absurdly incisive analytical skills. Thank you for all of the mentorship that contributed to what's inside this thesis; and, more importantly, all that's outside of it. Here's to many more projects.

I'd like to thank entirety of the Department of Psychology faculty for their mentorship since I began my undergraduate 4 years ago. Thank you all for your brilliance, insight, and facilitation of discussion (and my annoying questions). A huge thanks also to the administrative staff for everything, and Derek for being able to fix essentially anything.

To all my fellow postgrads - thank you for being great! In particular thank you to Aoife, Conor, Dylan, and Joanne (I alphabetised you...) - even though I had been here for 3 years previously, you guys added an entirely new function to Maynooth for me. Thank you for being some of the most intelligent, kind, and wonderful people I've ever met.
To Aoife, especially - thank you for your support, guidance, and brilliance (and culinary expertise!).

Finally, thank you to my family, especially my parents, Jason and Lisa. Thank you both for your continued support, both intellectually and financially. Thank you for dealing with my ever-narrowing fields of interest, my ever-increasing argumentativeness (blame academia), and my ever leftward-leaning political ideals. I would never have gotten anywhere near this point if it wasn't for the learning history you have provided me. Thank you for everything.

Abstract

The current research was focused on the investigation of the effectiveness of a new behaviour-analytic implicit psychological measure, the Function Acquisition Speed Test (FAST) in quantifying differential relatedness between experimental stimuli. Demonstration of the effectiveness of the measure in this regard would provide salient information to implicit measures researchers regarding the specific behavioural processes quantified by these measures, as well as providing stimulus equivalence researchers with a novel, continuous measure of differential relatedness. Experiment 1 (n = 62) investigated the effectiveness of the FAST in quantifying differential stimulus relatedness varied along inter-class parameters (i.e., a differing number of training iterations). Experiment 2 (n = 16) investigated FAST's efficacy in measuring differential intra-class stimulus relatedness, with subjects completing three FASTs with stimulus pairs of varying nodal distance based on training. The FAST showed effectiveness in both experiments in detecting differences in experimentally-varied relatedness. The findings from the current experiments suggest that the FAST is an effective measure of differential stimulus relatedness, providing an empirical basis for a behaviour-analytic stimulus relatedness account of effects seen in implicit measures. As well as this, the FAST shows promise in quantifying the emergence of stimulus relations during training procedures. Overcoming the conceptual and procedural opacity of other implicit measures, the FAST may be suggested as a functionally-understood, conceptually-coherent implicit psychological measure, with great potential utility in the quantification of stimulus relatedness.

Chapter 1

General Introduction

1.1 Introduction

The current research is focused on assessing the suitability of a behaviour-analytic test, the Function Acquisition Speed Test (FAST), as a measure of two dimensions of stimulus relatedness. Demonstrating the FAST's ability to quantify varying stimulus relatedness will provide stimulus equivalence researchers with a novel tool for inquiring into the complex dynamics of stimulus equivalence relations. In tandem with this, the ability to quantify relatedness will also provide support for the FAST as a measure of verbal learning histories, and may therefore be utilised by behaviour-analytic researchers in developing a functional understanding of how both cognitively and behaviourally oriented 'implicit' measures might work (i.e., what are the core processes and what exactly do they measure). This thesis will firstly provide an overview of stimulus equivalence research, as well as this concept of stimulus relatedness. Following this, an outline of the conceptual origins of the FAST will be detailed, as well as its procedural nuances. Finally, this thesis will argue for the conceptual amenability of the FAST as a measure of stimulus relatedness, and outline two experiments which seek to empirically test its utility in this regard.

1.2 A Brief History of Stimulus Equivalence

1.2.1 Early Behaviour-Analytic Approaches to Language

One of the first formal behavioural accounts of human language was that of B. F. Skinner (Skinner, 1957). Skinner suggested that verbal behaviour was the product of similar behavioural processes as other forms of behaviour; i.e., operant and respondent conditioning. In this Skinnerian conceptualisation, verbal behaviour is performed by a speaker, whose behaviour is reinforced or punished by a listener; this listener's contingent reinforcement or punishment is emitted as a consequence of their own conditioning by a verbal community (Hayes, Barnes-Holmes, & Roche, 2001). Although salient in extending the scope of behaviour analysis, Skinner’s account was criticised for the inter-organism dependency entailed within his definition. For instance, two topographically-identical behaviours may be seen as either verbal or non-verbal contingent on the mediator of the behaviour, which implicates the history of the speaker, rather than the listener, within its definition. This rendered Skinner’s definition non-amenable to a functional analysis (Hayes et al., 2001). As well as this, a widely-repeated criticism of Skinner’s view was that it was unable to account for the generativity of language, in particular the rapid rate which children come to learn language (Chomsky, 1959).

1.2.2 Sidman’s Paradigm

After the explication and criticism of Skinnerian verbal behaviour, a behavioural phenomenon known as stimulus equivalence was explored in a number of experiments by Sidman (2009). Stimulus equivalence is the phenomenon whereby two previously unrelated stimuli come to share a function through their relationship to a third stimulus. Using a match-to-sample (MTS) procedure, Sidman showed that when a sample stimulus (A) is established as a discriminative stimulus for the selection of a comparison stimulus (B) from a stimulus array containing B and some neutral stimulus (N), presenting B as a sample stimulus results in the selection of A from a comparison stimulus array consisting of A and N, in spite of a B-A directional discrimination having never been directly trained (Steele & Hayes, 1991). This phenomenon may be referred to as responding in accordance with symmetry. If a similar relationship is trained between B and a third stimulus (C), then when presented with A as a sample stimulus and a comparison array consisting of C and N, subjects will tend to select C in most cases (described by Sidman as responding in accordance with transitivity). This occurs in spite of A and C having never been presented together in any previous context. Strictly speaking, stimulus equivalence occurs only when C as sample discriminates the response of A as comparison, as equivalence requires the combination of symmetry and transitivity responding (Arntzen & Holth, 1997).

It should be noted that entailed within stimulus equivalence is the related phenomenon of transfer of function. Transfer of function describes the sharing of stimulus functions which occurs when stimuli are trained as equivalent (Gatch & Osborne, 1989). For instance, training a four member equivalence class, and then subsequently pairing one member of this class with a shock, leads to equivalent stimuli acquiring a similar function (Dougher, Augustson, Markham, Greenway, & Wulfert, 1994). Conversely, extinguishing a function of a particular stimulus also leads to extinction across equivalent members of that stimulus class. This phenomenon of transfer of function is a salient aspect of stimulus equivalence, in that stimuli equivalent by membership of a specific class may also be considered functionally equivalent (Dougher & Markham, 1994).

1.2.3 Stimulus Equivalence and Language

The phenomenon of stimulus equivalence is of particular interest to researchers in that it appears to be a uniquely human ability. Human children as young as four can exhibit stimulus equivalence across a variety of procedures and contexts (Devany, Hayes & Nelson, 1986; Egli, Joseph & Thompson, 1997; Pilgrim, Chambers & Galizio, 1995; Smeets, Barnes & Roche, 1997). By contrast, no non-human animal has been shown experimentally to respond in accordance with equivalence in current Sidmanian paradigms (Hayes, 1989; Urcuioli, Wasserman, & Zentall, 2013; though see Beums, Traets, De Houwer and Beckers (2017) for a discussion of differential stimulus functions, rather than lack of ability to form equivalence *per se*, as the cause of this). Given the apparently anthropocentric nature of this behaviour, it has been suggested that the operant capacity to form stimulus equivalence relations may be central to understanding human symbolic behaviour, as well as underlying the linguistic capacity of humans more generally (i.e., through the sharing of stimulus functions in equivalence classes; Dickins & Dickins, 2001). A number of studies have suggested that language and equivalence formation are at the very least related. For instance, Devany et al. (1986) showed that language-able children were able to form equivalence classes, while language-disabled children were not. As well as this, stimulus equivalence has also been utilised as a means of training language to children, both typically-developing and developmentally-disabled (Carr & Felce, 2000).

It has also been shown that linguistic learning histories can actively interfere with the formation of stimulus equivalence classes related to those learning histories. Watt, Keenan, Barnes, and Cairns (1991) demonstrated that the formation of stimulus equivalence relations was inhibited by prior verbal social learning. Specifically, Protestant and Catholic Northern Irish children could not derive equivalence between Protestant symbols and Catholic names in spite of training; however, this was not the case for Protestant English children (whose verbal learning histories likely did not entail the same Catholic-Protestant duality as Northern Irish children). In this sense, it is arguable that at the very least stimulus equivalence may provide an insight into the verbal learning history of individuals. The following section will provide a more detailed discussion of the process of stimulus equivalance.

1.3 Stimulus Equivalence and Differential Response Strengths

1.3.1 Stimulus Equivalence is a Behavioural Process

It should be regarded that stimulus equivalence is a compound behavioural phenomenon constituting of reflexivity (i.e., simple identity matching, e.g. A goes with A), symmetry, and transitivity. As such, stimulus equivalence should be understood in a truly Sidmanian sense as a behavioural *process*, rather than a dichotomous present/absent phenomenon *per se* (Sidman, 1960). For example, Pilgrim and Galizio (1990) found that, following the formation of baseline equivalence relations (i.e., A1-B1-C1 and A2-B2-C2, where alphanumerics represent some arbitrary stimulus, such as nonsense words or symbols), training A-C conditional discriminations in opposition to those derived at baseline (i.e., training A1-C2 and A2-C1) led to changes in symmetry relations from baseline, but not to changes in transitivity relations. Other studies have shown similar divergence between symmetry and transitivity relations following baseline equivalence formation (see Pilgrim & Galizio, 2000). As well as this, the presence of all three of these components does not necessarily lead to the immediate emergence of stimulus equivalence. In many instances, stimulus equivalence is only demonstrable with multiple iterations of testing trials (e.g. Spradlin, Cotter, & Braxley, 1973).

The evidence for the necessity of multiple testing trials before stimulus equivalence is observed implies that the accuracy of responses in accordance with equivalence improves with a greater number of testing trials. Even when equivalence responses occur with 100% accuracy, other dimensions of the process are susceptible to change. In particular, Imam (2001) noted that even after accuracy scores reach a 100% level, responses latencies on equivalence testing trials tend to shorten as the number of stimulus presentations increased. Stimulus equivalence is thus not only a compound behavioural phenomenon; the degree of stimulus equivalence responses also tend to strengthen (in terms of accuracy and response latency) with increased training and opportunities to respond[1].

1.3.2 Resistance to Change

The above mentioned strengthening of responses may be understood in terms of the conceptual framework of resistance to change. Resistance to change may be understood as

---

[1] Readers should note that issue may be taken with the ontological notion of response strength, in that “response strength” may be seen as reliant on an implicit notion of a ‘store’ of reinforcement. While this question is beyond the scope of the current thesis, see Shahan (2017) for a detailed d.iscussion of this conceptual issue.

consisting of the resistance of a specific response elicited by a subject to behavioural change in the presence of contingencies at odds with the response (Nevin, 1988). Behaviour is understood as more resistant if it is emitted at a high rate in the presence of a salient discriminative stimulus (i.e., as a consequence of a greater history of reinforcement), and if it persists for a long period of time in spite of competing contingencies (Nevin & Grace, 2000). Resistance to change has been studied experimentally. For instance, Nevin (1992) found that differentiating prior ratios of reinforcement for responses led to differential resistances to change for those responses in pigeons, with a higher reinforcement ratio leading to greater resistance to behavioural change. Similarly, Mace et al. (1990) found that mentally-disabled adults were less inclined to be distracted from a sorting task by alternative stimuli when the rate of reinforcement was higher (60-s versus 240-s). Resistance to change has also proven robust in applied settings, with the concept providing utility in understanding treatment relapse following behavioural intervention (Pritchard, Hoerger, & Mace, 2014), compliance to adult requests in a number of contexts (Radley & Dart, 2016), academic task-switching (Lee, 2006), and general academic engagement (Martens, Lochner, & Kelly, 1992).

It is arguable that stimulus equivalence responses in the MTS may be understood in terms of resistance to change. That is (as discussed in 1.3.1), as more opportunities for training and testing are provided, subjects become more accurate and quick (i.e., their rate of responding increases) at responding correctly to equivalent stimuli. As well as this, the Watt et al. (1991) study may provide support for this perspective. That is, in that study, subjects did not respond in accordance with new contingencies for equivalence (i.e., Catholic symbols with Protestant names) due to their previous and more extended learning histories with those stimuli (i.e., resistance to behavioural change was strong). Further, Dickins, Bentall, & Smith (1993) found that stimuli trained in equivalence classes were less likely to elicit equivalence responses from subjects if those stimuli were also trained as being in some way incongruous (although still having been trained as equivalent). It may be inferred from this latter study that manipulating the relationship between directly-trained stimuli can subsequently affect the strength of the equivalence responses which emerge from their pairing (e.g., differential strength of A-B and B-C responding leads to commensurate changes in the strength of A-C responses). In effect, resistance to change appears to be a feature of stimulus equivalence and should be, measureable where reinforcement histories for such responding have been manipulated.

1.3.3 Stimulus Relatedness

A related phenomenon to resistance to change which is specific to stimulus equivalence is that of stimulus relatedness. It should be noted that stimulus relatedness as a concept is relatively ill-defined within the literature. Indeed, while reference to the concept is seen in a wealth of behavioural manuscripts (a cursory Google Scholar search yields 121 papers which use the exact phrase ‘stimulus relatedness’), no formal operationalisation of the concept is consistent across studies. However, given the various uses of the term, relatedness may be broadly understood at this point in time as the degree to which stimuli are related to one another, which is most intuitively measured by the probability of their being related in a matching context. Indeed, where it has been operationalised, relatedness is typically measured in terms of the likelihood of selecting a comparison stimulus in the presence of a given sample (Moss-Lourenco & Fields, 2011). The experimental manipulation of relatedness is generally operationalised through variations in the nodal distance of equivalence class members. That is, in a linear five-member equivalence class (A-B-C-D-E, with A-B, B-C, C-D, and D-E reinforced), A-C stimulus equivalence will exhibit greater relatedness than A-D, and A-D will exhibit greater relatedness than A-E. As such, stimulus relatedness is thought to increase as nodal distance decreases (Fields & Moss, 2007).

While stimulus relatedness is typically manipulated in equivalence literature via nodal distance, this is not necessarily the exclusive parameter by which it may be understood. If relatedness is typically measured in terms of probability of stimulus selection given a sample stimulus, then paradigms which are associated with the concept of resistance to behavioural change should also fall within the operationalisation of relatedness. That is, class consistent responding becoming more likely necessarily entails a reduction in the likelihood of class-inconsistent responses. In effect, in bringing together the concepts of stimulus relatedness and resistance to change, we might expect that incrementally-larger nodal distances would lead to commensurately lowered resistance to change. Interestingly, such a phenomenon has been noted in the literature. Fields, Landon-Jimenez, Buffington, & Adams (1995) found that the likelihood of the stimulus function trained to an A stimulus transferring to another given class member varied inversely with the nodal distance between the relevant stimuli. It has also been found that the stability of transferred stimulus functions (a key element in degree of resistance to change) varies in accordance with nodal distance (Rehfeldt & Dymond, 2005). Fields and Moss (2007) summarised a large body of evidence exhibiting similar nodal distance effects, and proposed that stimuli in equivalence classes acquire two functions: one based on class membership (referred to herein as ‘inter-class’ responding), and the other based on nodal distance (referred to herein as ‘intra-class’ responding; see Figure X). All stimuli in an equivalence class are functionally equivalent when an inter-class response is required. However, stimuli are not functionally equivalent in terms of intra-class responding, and vary in accordance with the nodal distance between relations. From this perspective, it may be coherent to suggest that resistance to change research has typically dealt with inter-class stimulus relatedness effects, while nodal distance research is reflective of intra-class stimulus relatedness effects. From the overlapping qualities which parameterise both of these phenomena (probability of stimulus selection and relative resistance to change/persistence of function), it may therefore be suggested that stimulus relatedness may be manipulated not only through differential nodal distancing, but also through manipulations of resistance to change (e.g., differential iterations of MTS training in specific equivalence class elements). One such paradigm which utilised this approach was that of Bortoloti, Rodrigues, Cortez, Pimentel, and de Rose (2013), which found that the probability of selecting specific class-consistent stimuli (i.e., stimulus relatedness) increased in accordance with overtraining of stimulus class members.

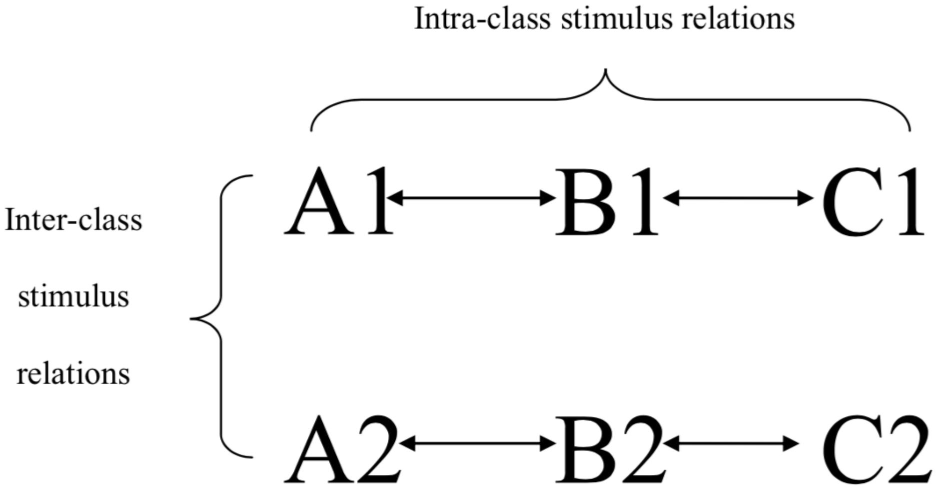


Figure 1 An illustration of the difference between inter- and intra-class stimulus relations. A1-B1-C1 represents an equivalence class, and A2-B2-C2 represents another equivalence class.

1.4 Quantifying Stimulus Relatedness

There are a number of means by which stimulus relatedness is currently quantified in the experimental literature on stimulus equivalence. Most methods require the use of a MTS-style test procedure in some capacity to assess relatedness. For instance, Moss-Lourenco and

Fields (2011) detailed two common approaches to quantifying relatedness: post-class formation preference tests, and dual-option function transfer tests. Experimental post-class formation preference tests involve the presentation of various members of a trained equivalence class as sample and comparisons, and observing whether responses are affected by the manipulated relatedness value (i.e., nodal distance). Dual-option function transfer tests, by contrast, involve training functions to some individual stimuli in an equivalence class, and observing how functions transfer across the class in accordance with varying relatedness. For instance, Fields and Watanabe-Rose (2008) used linear MTS training to establish two 6-member equivalence classes (A1/2-B1/2-C1/2-D1/2-E1/2-F1/2). The C stimuli were trained to discriminative three presses of the ‘J’ computer key, while the D stimuli were trained to discriminative seven presses of the same key. C’s function transferred along the B and A stimuli, while D’s function transferred along the E and F stimuli. Both of these instances of transfer of function varied in strength in each stimulus in accordance with nodal distance effects, which provided evidence for differential stimulus relatedness across nodes. One additional (though less common) approach to assessing stimulus relatedness has been to use a post-class formation test with two metrics: response accuracy and response times (Bentall, Jones, & Dickins, 1999). Response latencies tend to increase as a function of nodal distance, and as such, shortened response latencies are assumed to signify greater stimulus relatedness (Spencer & Chase, 1996).

Readers should note that, as previously mentioned, stimulus relatedness is typically manipulated via differential nodal distance. However, as discussed, stimulus relatedness may also be manipulated through inter-class differences. As such, while the aforementioned approaches to quantification are appropriate for intra-class measurement, they may not be appropriate for inter-class stimulus relatedness quantification. For example, experimentally manipulating inter-class stimulus relatedness could involve providing multiple iterations of

MTS training to subjects. If testing for stimulus relatedness was in a similar format to the training (i.e., match-to-sample, as in all of the above methods), then response times in the test may be quicker with more training simply due to procedural practice, rather than due to greater relatedness between the stimuli per se (see for example Pashler & Baylis, 1991a, 1991b). Indeed, the earlier mentioned Bentall et al. (1999) found that response latencies generally decreased as testing progressed in their procedure. While it could be argued that this issue may be ameliorated by opting not to measure response times and focus solely on the accuracy of responses (as is the case in post-class formation preference tests and dual-option function transfer tests), excluding response times would eliminate a key metric which is known to vary in equivalence classes as a function other variables (Spencer & Chase, 1996). Given that a comprehensive measure of stimulus relatedness should be able to effectively quantify both intra- and inter-class differences, current methods for assessing relatedness are therefore at an impasse. What is needed is a measure which can quantify differential relatedness irrespective of whether it varies along inter- or intra-class parameters. Such a measure would also require amenability to analyses of both response times and accuracy of responding, or some amalgam of the two. It may be argued that a new behaviour-analytic measure, the Function Acquisition Speed Test, offers such a methodology.

The Function Acquisition Speed Test (FAST) involves the training of two functional responses towards two pairs of stimuli, with pairings juxtaposed across two blocks. It is assumed that observed differences in the fluency of responses on one block in comparison to the other are indicative of differences in the contiguity of the paired stimuli relative to the learning history of the subject (Cartwright, Roche, Gogarty, O’Reilly, & Stewart, 2016). Readers should note that response fluency is the primary metric of interest within the FAST. Fluency refers to the amalgamation of response time and accuracy as some singular metric (Binder, 1996). Further, the procedural features of the FAST are different to those of the matching-to-sample. As such, the FAST may be purported as a viable alternative to the MTS in quantifying stimulus relatedness, in that it utilises a robust metric and avoids the potential pitfalls of MTS-style testing procedures in measuring experimentally-varied inter-class relatedness. Before full discussion of the FAST can be undertaken, however, it is salient for the reader to have an appreciation for the original research context from which the FAST was developed.

1.5 Background of the Function Acquisition Speed Test

1.5.1 Implicit Measures and the Implicit Association Test

In the 1990s within social psychology, typically-used explicit-response methods for assessing ‘attitudes’ of individuals received heavy criticism due to the tendency for responses to influenced by social-norms, leading to responding based on social desirability (rather than truthful responding; see Moorman and Podsakoff (1992), for example). Greenwald and Banaji (1995), by contrast, introduced the construct of “implicit attitudes” as a purported alternative to understanding behaviour in entirely functional terms. “Implicit attitudes” were proposed as being the primary influencer of behaviour of individuals, but are also considered to be inaccessible to those individuals. Greenwald, McGhee, and Schwartz (1998) explicated upon this construct further, and outlined a new measure for quantifying these “implicit attitudes”: the Implicit Association Test (IAT; see Figure 2).

The IAT involves subjects responding on a computer keyboard based on rules that instruct a specific response to each of four stimulus types (i.e., from one of two target stimulus classes or one of two attribute stimulus classes) that are presented individually on a computer screen. Subjects are instructed to respond as quickly and as accurately as possible, though notably, no response window is enforced within the procedure itself. Instead, response times which are over 3000ms are rounded down to 3000ms post-hoc.

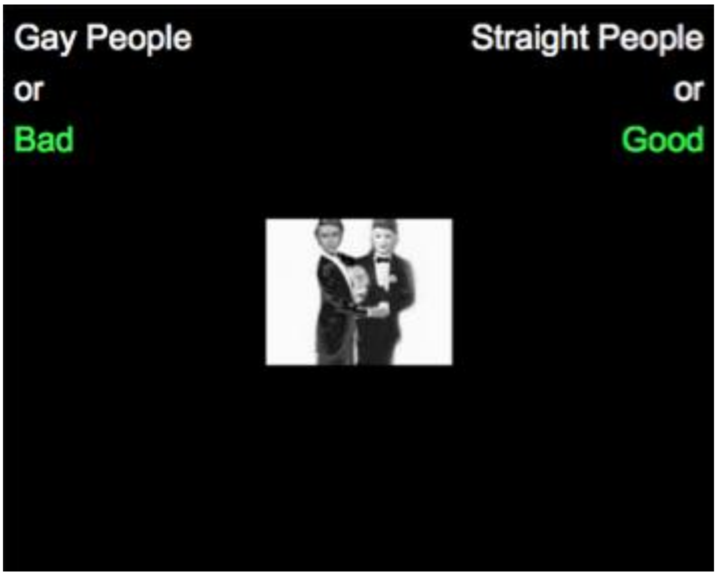


Figure 2 Screenshot of a trial from a IAT to measure implicit attitudes towards gay individuals. This trial is from a block consistent with an anti-gay bias, where subjects are required to respond with the ‘d’ key (i.e., leftward) when either a gay person image or bad word stimulus appears on-screen, and respond with the ‘k’ key (i.e., rightward) when a straight person image or good word stimulus appears on-screen. An inconsistent block would involve an orthogonal configuration of responses.

Likewise, response times below 400ms are rounded up to 400ms.The precise test format is as follows. Two 20-trial practice blocks for single-target categorization (e.g., block 1: insect/flower, block 2: good/bad). A 20-trial block (3) then follows with both target and attribute stimuli present (e.g., insect-good, flower-bad); this block is immediately followed by a 40-trial block (4) with identical stimuli to (3). The next block (5) returns to a single-single target categorization task with the category sides switched, for 20 trials (e.g., flower/insect). Blocks 6 and 7 replicate the structure of blocks 3 and 4, but with alternated stimulus pairs (e.g., insect-bad, flower-good). The subject must respond with a key press (typically the ‘d’ or ‘k’ key) as instructed in response to each stimulus. Feedback is provided only on error trials, with positive reinforcement never delivered at all. Notably, only blocks 4 and 7 are considered ‘critical’ blocks (i.e., the blocks which are scored). A difference score (*d*-score) based on the variance in response latencies between the two critical blocks is used to infer the nature of the subject’s “implicit attitude” regarding the tested stimuli. This scoring system is based on the reasoning that faster responses during one block compared to the other indicate class compatibilities (e.g., insects-bad / flowers-good) that are congruous with the subject’s personal cognitive biases, which in turn can be used as a proxy for ‘attitudes’ (Greenwald & Banaji, 1995).

1.5.2 What Does the IAT Measure?

It should be noted that the creators of the IAT have been relatively conservative in attempting to precisely define the ‘cognitive biases’ the IAT is purported to measure (Greenwald & Nosek, 2008). This lack of specificity has led to IAT research being of particularly wide scope, with the test being purported as predictive of phenomena as diverse as self-esteem, anxiety, future depressive symptoms, suicidality, self-injurious thoughts, dietary preference, pedophilia, alcohol addiction, risk-taking behaviour, future weight gain, diagnostic behaviour of physicians, voting behaviour, positive environmental behaviour, smoking behaviour, responses towards aggravation, likelihood of nurses changing jobs, and opposition to former US president Barack Obama’s health reform plans, to name but a few (Dal Cin, Gibson, Zanna, Shumate & Fong, 2007; Egloff & Schmukle, 2002; Franck, Raedt & Houwer, 2007; Gray, Brown, MacCulloch, Smith & Snowden, 2005; Green et al., 2007; Greenwald & Farnham, 2000; Greenwald, Smith, Sriram, Bar-Anan & Nosek, 2009; Jajodia & Earleywine, 2003; Knowles, Lowery & Schaumberg, 2010; Mauss, Evers, Wilhelm & Gross, 2006; Molesworth & Chang, 2009; Nederkoorn, Houben, Hofmann, Roefs & Jansen, 2010; Nock et al., 2010; Nock & Banaji, 2007; Richetin, Perugini, Prestwich & O'Gorman, 2007; Schultz, Shriver, Tabanico & Khazian, 2004; von Hippel, Brener & von Hippel, 2008). Somewhat paradoxically, the purported inferential generality of the implicit attitude construct is contrasted by the highly-specific conditions under which it is claimed to be measurable. That is, by subtracting the mean response latency of the 4th IAT test block from that of the 7th IAT test block (or vice verse, depending on which is the ‘consistent’ block) and dividing this value by the standard deviation of response times across both critical blocks. It should be noted that the use of this d-score and test format is based primarily on a single paper from Greenwald, Nosek, and Banaji (2003), which compared the effectiveness of 10 different metrics on 7 different criteria (correlation with explicit measures, correlation with average latency, internal consistency, order effect correction, IAT experience correction, effect size, and implicit-explicit path in a confirmatory factor analysis; though see Richetin, Costantini, Perugini, & Schönbrodt, 2015, for a more recent analysis). Of particular interest is the criterion of correlation with explicit measures. While explicit-implicit correlation is taken as evidence for the IAT’s effectiveness in some studies (e.g., Banse, Seise & Zerbes, 2001; Egloff & Schmukle, 2002; McConnell & Leibold, 2001), in many other IAT studies, a lack of correlation with an explicit measure is also taken for evidential value (e.g., Stull, McGrew, Salyers & Ashburn-Nardo, 2013; Greenwald et al., 1998).

Aside from the argument that the above practice may suggest that methods of validating the IAT are unfalsifiable, it is curious that the convergence of the implicit measure with explicit measures has been used to justify its applicability, given that the implicit attitude construct ought to in some way diverge from explicit scores (Greenwald et al., 1998). While it is arguable that some degree of convergence should also be seen (cf. Cronbach & Meehl, 1955), it still remains unclear what degree of correspondence between explicit and implicit measures should be seen, and what precise source of any such correspondence would be. The conceptual inconsistency in this reasoning may be argued as symptomatic of the non-specific nature of the implicit attitude construct. However, in spite of the aforementioned conceptual issues which accompany the social-cognitive theoretical interpretation of these effects, the IAT appears to capture some aspects of class-based responding that is distinct from the explicit verbal responding typically recorded using explicit paper and pencil measures of “attitudes”. In this sense, the behavioural phenomenon captured by the IAT is of interest. Indeed, behaviour-analytic researchers have attempted to account for the effects noted on implicit measures whilst simultaneously avoiding the issues of social-cognitive theory. Two main behaviour-analytic measures have been developed in this regard: the Implicit Relational Assessment Procedure and the Function Acquisition Speed Test.

1.5.3 The Implicit Relational Assessment Procedure

The Implicit Relational Assessment Procedure (IRAP) originates from the theoretical position of Relational Frame Theory (RFT; Hayes, et al., 2001). RFT is a modern behaviour-analytic account of human language and cognition which, in brief, suggests that human language is based on the ability to derived arbitrary relations between stimuli based on their relations to other stimuli (Dymond & Roche, 2013). That is, RFT considers the underlying process involved in human language and cognition to be arbitrarily-applicable relational operants. Under RFT, stimulus equivalence responses are a form of the relational operant of ‘sameness’; a relational operant of ‘opposition’ would involve (for example) training that A is opposite to B, and training that B is opposite to C. From this, subjects should derive that A is the same as C, and that C is the same as A (as the opposite of ‘opposite of A’ is the same as A; Hayes et al., 2001). RFT extends these relational operants to include a number of other forms of responding, for example temporal relations, hierarchical relations, and deictic relations (Colbert, Dobutowitsch, Roche, & Brophy, 2017). From this perspective, effects in implicit measures such as the IAT are the consequence of complex verbal learning histories towards the relevant stimuli. Extending from this theoretical framework, the Relational Elaboration and Coherence (REC; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010) model suggests that when responding in implicit procedures, subjects tend to elicit two responses: a first response, known as a brief immediate relational response (BIRR), is elicited based on the subject’s learning history with the relevant stimuli, while a second, more elaborated response is elicited based on the specific instruction in the procedure in accordance to which the subject is attempting to respond (Barnes-Holmes et al., 2010). It is assumed that contiguity between these two responses will facilitate speed of the responding in the procedure, while incompatibility of these responses will lead to slower responding. As such, the slower test block in an implicit measure is assumed to be incongruous to the subject’s learning history.

RFT and the REC model are the conceptual foundations upon which the IRAP implicit measure was developed. The IRAP involves the presentation of a pair of stimuli on-screen (e.g., black/white, good/bad), and requires subjects to respond to these pairings in accordance with a specific relational rule (e.g., SAME, SIMILAR, MORE-THAN; see Barnes-Holmes et al., 2006). The relevant rule may be presented on-screen with the stimuli, and subjects must respond based on a key-press (e.g., ‘d’ for ‘TRUE’, ‘k’ for ‘FALSE’). Alternatively, an IRAP may require subjects to directly specify the relation using this key-press (e.g., pressing the ‘d’ key for ‘SAME’ and the ‘k’ key for ‘NOT SAME/DIFFERENT’). In this sense, the IRAP is distinct from the IAT, in that it directly specifies the relationship between experimental stimuli. Similar to the IAT, stimulus pairs are juxtaposed across critical blocks in the IRAP. For instance, an IRAP testing for differences in relating dog and carrot stimuli to animal or vegetable stimuli would involve one critical block wherein subjects would be required to respond to the constantly present comparison stimuli (i.e., contextual cues) ’SAME’ when presented with a dog/animal pair or a carrot/vegetable pair, and ‘DIFFERENT’ when presented with a dog/vegetable or carrot/animal pair. A separate block would involve the converse of this configuration.

As with the IAT, the IRAP also consists of a number of non-critical blocks. Though IRAP configurations vary slightly across studies, the standard IRAP structure is as follows (Hussey, 2017). An initial set of practice blocks pairs is presented, using the relevant experimental stimuli. ‘Practice block pairs’ refers to the fact that two blocks are presented: one such block involves a rule consistent with an assumed repertoire, while the other involves a rule inconsistent with the assumed repertoire. Subjects are required to achieve minimum criteria of 80% accuracy and median response times less than 2000ms on both of the practice block pairs in order to proceed to the testing phase. If subjects fail to reach these criteria, then the practice block pairs are recycled. If the subject fails to reach criteria of the practice block pairs after 4 cycles, then the procedure is terminated. Once practice criteria are met, the subject moves on to the test phase, which consists of three iterations of each experimental block (i.e., consistent and inconsistent). Each block throughout the procedure consists of 32 trials (i.e., 8 iterations of each category-stimulus combination). Mean response latencies (with response times over 10s and under 400ms excluded) are compared across the test blocks (i.e., the mean of the three consistent test blocks versus the mean of the three inconsistent test blocks), and the difference in these latencies divided by their pooled standard deviation (represented by a score known as the *D*-IRAP) is taken as a measure of differences in the verbal learning history (i.e., the IRAP effect; see Table 1).

1.5.4 Procedural and Conceptual Issues with the IRAP

Although the IRAP overcomes the conceptual ambiguity of the IAT through its reliance on a functional account of implicit effects, as well as improves upon the specificity of the IAT procedure by allowing for specific relational operants to be targeted, there are a number of procedural and conceptual issues which arise from its use. For instance, the REC model relies on the notion of two verbal responses being emitted by the subject during the IRAP procedure. In particular, the initial Brief and Immediate Relational Response emitted is

Table 1. Side-by-side comparison of the procedural features of the IAT and IRAP.

	Procedure	
Feature	**IAT**	**IRAP**
Purported to measure	Mental associations	Verbal learning histories (REC model)
Feedback	Incorrect responses only; subject must correct response while RT recording continues	Incorrect responses only; subject must correct response while RT recording continues
Format	Fixed number of blocks; 7 different blocks; 20 or 40 trials per block; 4th and 7th blocks are critical	Variable number of blocks (depending on how quickly practice criteria are reached, i.e., >= 80% correct, median RT <= 2000ms); 1-4 practice block pairs, 3 test block pairs (i.e., 3 blocks with consistent rule, 3 blocks with inconsistent rule); final 3 block pairs are critical
Response window	No response window in procedure, values are truncated or removed post-hoc	No response window in procedure, red '!' shows after 2000ms, values are truncated or removed post-hoc
Metric/scoring	D1 score: (Mean Incon RT – Mean Con RT) / Pooled RT_{SD}	D-IRAP score: (Mean Incon RT – Mean Con RT) / Pooled RT_{SD}

conjectured to be the salient factor in differentiating response times across blocks in the IRAP. However, it is unclear as to how this specific response could be investigated empirically without resorting to tautological fallacies in reasoning (for instance, a paradigm that attempts to manipulate this BIRR could be achieved through experimental control of relations between stimuli, but the IRAP would be required for the quantification of this change, which thus leads to tautology, given that a number of conflicting explanations could also be put forward with equal veracity). Hence, it is arguable that the notion of the BIRR is unfalsifiable, and thus lacks the functional utility to make meaningful inferences about the source of implicit measures (a lá the scientific philosophy of Popper, 1934).

Aside from the above-mentioned conceptual ambiguity, a number of issues may also be taken with a number of procedural aspects of the IRAP. For example, there is no empirical justification in the published IRAP literature for the exclusive presentation of feedback on incorrect responses, and no published literature on the potential confounding effects of feedback on subsequent trials. At the very least, this aspect of the procedure is unexplored, and could potentially interfere or influence effects noted in the procedure. As well as this, the use of the D-IRAP score may be criticized, and in particular the removal of response times above 10 seconds and less than 400ms. Specifically, it is arguable that a 10 second cut off for response times is an arbitrary value with little empirical grounding to justify its use. Also notable is the lack of explicit response window to enhance the fluency which is critical to the emission of the hypothetical BIRRs, although a red ‘!’ is displayed after 2000ms to prompt an increase in speed. Importantly, however, this prompt does not force the subject to make a selection, and their elongated response time is still included in all data analyses so long as it does not exceed 10s. This is a curiously weak form of behavioral control.

It is of interest, further, to note that although the IRAP employs accuracy criteria in its practice blocks, accuracy is not included in the quantification of the D-IRAP score. This is particularly interesting given that the traditional metric of interest in behaviour analysis is primarily that of accuracy (Pierce & Cheney, 2013). It may be argued that the IRAP includes accuracy in its D-IRAP metric via the procedural necessitation of a corrective response following an erroneous response, but it is a curious and unorthodox procedure to attempt to increase a correct response rate by only punishing incorrect responses. The anachronistic procedure of only punishing incorrect responses appears to have been inherited from the IAT literature in whole cloth and has made its way into a purportedly behavior-analytic test for stimulus relations, but without the attendant bottom-up analyses available to justify the practice. In addition to this, the use of the D-IRAP score involves the implicit commensuration of response time with accuracy. That is, subjects are required to correct erroneous responses, and the time taken to correct the erroneous response is also included in the overall response time value for that trial. As such, greater inaccuracy is ‘encoded’ into the D-IRAP score via this additional time. The use of this penalizing feature arguably involves the assumption that the time taken to correct an erroneous response is in some form a fitting analogue for accuracy. This in turn would then imply the existence of some generally-applicable transformative function to translate accuracy to response time (i.e., by assuming accuracy and response times are commensurable). However, given that response times and accuracy are at least in some dimension conditionally-dependent, such a transformation is not possible (Bolsinova, de Boeck, & Tijmstra, 2016; Bolsinova, Tijmstra, Molenaar, & de Boeck, 2017). In this sense, the D-IRAP measure fails to account for accuracy of responses in a meaningful way, and the potential explanation of the analogue of requiring corrective responses is not a conceptually-coherent approach to quantifying accuracy.

On a related point, it is curious that the IRAP attempts to raise accuracy in its criterion for training blocks, but then excludes any explicit analysis of accuracy in its indexing of scores on critical blocks. A reasonable explanation for this may be that stabilizing the accuracy of subjects leads to a greater potency in the analysis of response times. That is, in reducing the variance in the accuracy of subjects’ responses, the difference in the difficulty of trials will be more greatly pronounced in terms of trial accuracy (Cauraugh, 1990). This would be coherent to the finding within stimulus equivalence research that differences in response latencies are only sufficiently pronounced when accuracy has stabilized (Spencer & Chase, 1996). However, such an argument has never been explicated within the IRAP literature, nor empirically-investigated within the context of the IRAP procedure. As such, there is no published justification for the disconnect between the implementation of an accuracy criterion in the IRAP training blocks and the subsequent exclusive metrical focus on response times in the critical IRAP blocks. From the above criticisms, it is evident that while the IRAP ameliorates some of the conceptual issues surrounding the IAT, it also raises new problems.

1.6 The Function Acquisition Speed Test

The Function Acquisition Speed Test (FAST) represents an alternative attempt within behaviour analysis to inquire into the effects noted in implicit measures. The FAST originated as an attempt to develop a behavioural version of the IAT using the logic of stimulus equivalence and the now widely cited Watt et al (1991) effect (seeRoche, Ruiz, O'Riodran, & Hand, 2005). As such, the FAST is similar to the IRAP in the sense that effects noted within the procedure are conceptualised in the context of the verbal learning history of the subject. However, the FAST avoids conceptual issues which are associated with the assertion of the REC model, instead understanding differences in verbal learning histories through the well-understood (and previously discussed) principle of resistance to change (Cartwright, Roche, & O'Reilly 2015). It should be noted, therefore, that while the FAST fits the criteria to be described by social-cognitive researchers as an implicit measure (see De Houwer, Teige-Mocigemba, Spruyt, and Moors, 2009), the FAST's behaviour-analytic conceptual foundations consider constructs such as 'implicit attitudes' as merely the reification of hypothetical variables (see MacCorquodale & Meehl, 1948; Skinner, 1971). As such, the reader should understand that reference to the term 'implicit' here is for heuristic convenience only. The FAST has been developed through a data-driven research approach, and as such also arguably avoids the procedural pitfalls noted in both the IAT and the IRAP (see O'Reilly, Roche, Ruiz, Tyndall, & Gavin, 2012). In what follows, the FAST's procedure and scoring will be outlined in order to give the reader a full appreciation of the key differences between the FAST and the other formerly-discussed implicit measures.

1.6.1 The FAST Procedure

The current format of the FAST procedure (Cartwright et al., 2016) consists of two fifty-trial blocks of stimulus discrimination tasks. That is, unlike the IAT and IRAP, there are no non-critical blocks in the FAST. Further unlike the IAT and the IRAP, which involve the categorisation of stimuli, the FAST requires subjects to form two functional response classes for pairs of stimuli, with the specific pairings differing across two blocks. That is, both FAST blocks present the same stimuli, but blocks differ in terms of the reinforcement contingencies applied. More specifically, in each block, two functional response classes are established simultaneously, whereby exemplars from two distinct stimulus categories are discriminative for a positional keyboard response in each case. Like other implicit measures, it is assumed that one block's contingencies are consistent with the subject's learning history, while the other block's contingencies are inconsistent. For instance, in one block, subjects may be required to press the left-hand 'z' key when 'flower' or 'good' exemplars are presented on-screen, and the right-hand 'm' key when 'insect' or 'bad' stimuli are presented. In the other block, 'flower' and 'bad' stimuli would share a functional response, as would 'insect' and 'good' stimuli. Based on the principle of resistance to change, subjects should find responding in accordance with contingencies inconsistent with their learning histories more difficult than responding to contingencies consistent with this history.

It is important to note that in the FAST no rules are presented to the subject at any stage. Behaviour is shaped by reinforcement and punishment ('correct' or 'wrong' feedback) alone. As well as this, a 3 second response window is enforced. If the subject does not respond within this 3 second window, then the feedback 'wrong' is presented and the trial ends. In this sense, subjects' responses are shaped to be not only accurate, but also relatively quick. The reader may note a significant difference between the previously-discussed implicit measures and the FAST, in that corrective feedback is presented for both incorrect *and* correct trials. In this sense, the FAST overcomes the issue of providing feedback only on incorrect trials noted in other implicit measures. As well as this, the presence of a response window avoids the necessity of post-hoc truncation of scores notable in other implicit measures. Further still, erroneous responses result in the presentation of corrective feedback and the termination of the trial. That is, subjects are not required to correct their response in order to proceed to the next trial, which therefore avoids confusion regarding the purpose of feedback and response correction; to directly enhance accuracy, to indirectly enhance speed, or opaquely, to influence both in combination.

1.6.2 Current FAST Scoring

Like other implicit measures, scores on the FAST are calculated based on differences between responses on a consistent and inconsistent block. While the IAT and IRAP typically rely on *D* transformations of response times in the interpretation of effects, the FAST instead utilises a measure which amalgamates both response time and accuracy (i.e., fluency; Cartwright et al., 2016). Specifically, the cumulative learning curves recorded for each block are fitted with regression lines (herein referred to as the 'block-slope'), which are then contrasted. The size and direction of difference between block-slope scores is taken to indicate the degree to which one pair of response classes was acquired more easily than the other, referred to as the FAST effect. This provides an indirect or "implicit" measure of the stimulus relation configurations that existed pre-experimentally in the repertoire of the subject (see O'Reilly et al., 2015, for a conceptual analysis of this effect, and Cartwright et al., 2016 for a recent empirical analysis using real-world stimuli). This metric is arguably more comprehensive than the *D* score, in that it accounts for both accuracy and response times, and avoids the conceptual issues associated with the use of post-hoc manipulation of the data (see Ridgeway, Roche, Gavin, & Ruiz, 2010).

1.6.3 Variations of the FAST Format

Importantly, elements of the FAST procedure are not prescriptive. There have been a number of procedural and metrical variations on the FAST in different research contexts. For instance, the earliest studies using the FAST methodology simply measured the number of correct responses on consistent versus inconsistent blocks (Roche et al., 2005; Gavin, Roche, & Ruiz, 2008). Other studies (e.g., O'Reilly et al., 2012) using the FAST employed varying block lengths (where blocks terminated only when the subject achieved a criterion of 10 responses in-a-row correct). These studies also included initial baseline blocks with neutral stimuli in order to assess the baseline acquisition rates of subjects. The use of these baseline blocks enables a 'strength of relation' (SoR) index to be calculated. The SoR involves subtracting the count of trials on the consistent block from the count of trials for the inconsistent block, and dividing this value by the mean number of trials taken in the baseline blocks. These variants outlined above may provide more appropriate alternatives to the FAST method outlined in 1.6.1 and 1.6.2 if researchers are interested primarily in the accuracy of responding. In effect, the relative amorphousness of the FAST procedure is advantageous over other implicit measures which are not so readily malleable to the needs of the experimenter and appear to have become reified in the same way in which psychometric tests become co-defined by the constructs they are purported to measure (see table 2 for an overall comparison between the FAST and the other implicit measures discussed in the current work).

It should be noted that a variable block length is typically not amenable to analyses of mean response times, given the fact that achieving criterion in a lesser number of trials paradoxically penalises mean response times. That is, consider a subject whose first three response times on both the consistent and inconsistent block are 3000ms, 2700ms, and 2300ms respectively (i.e., long response times). From then on, the subject responds at the same rate on both blocks, with an average response time thereafter of 1000ms per trial. On the consistent block, the subject reaches criterion after 10 trials; on the inconsistent block, the subject takes 30 trials to achieve this.

Table 2. Side-by-side comparison of the procedural features of the IAT, IRAP, and FAST.

	Procedure		
Feature	**IAT**	**IRAP**	**FAST**
Purported to measure	Mental associations	Verbal learning histories (REC model)	Verbal learning histories (class compatibilities and resistance to change)
Feedback	Incorrect responses only; subject must correct response while RT recording continues	Incorrect responses only; subject must correct response while RT recording continues	All responses
Typical format	Fixed number of blocks; 7 different blocks; 20 or 40 trials per block; 4th and 7th blocks are critical	Variable number of blocks (depending on how quickly practice criteria are reached, i.e., >= 80% correct, median RT <= 2000ms); 1-4 practice block pairs, 3 test block pairs (i.e., 3 blocks with consistent rule, 3 blocks with inconsistent rule); final 3 block pairs are critical	Fixed number of blocks; one consistent, one inconsistent; 50 trials per block

Response window	No response window in procedure, values are truncated or removed post-hoc	No response window in procedure, red '!' shows after 2000ms, values are truncated or removed post-hoc	Response window of 3000ms, failure to respond results in 'incorrect' feedback
Typical metric/scoring	D1 score: (Mean Incon RT – Mean Con RT) / Pooled RT_{SD}	D-IRAP score: (Mean Incon RT – Mean Con RT) / Pooled RT_{SD}	FAST effect: Difference in the learning curve slopes across blocks

In spite of achieving criterion in much fewer trials on the consistent block compared to the inconsistent block, the subject's mean response time would be 1500ms in the consistent block, while the subject's mean response time in the inconsistent block would be 1166.67ms. This is due to the fact that a greater number of trials allows for the normalisation of the data, which in turn diminishes the impact of individual longer response times in the overall mean score. In this sense, FAST procedures whose blocks vary based on the criterion of 'correct-in-a-row' are not conducive to meaningful inferences about subject's response times.

1.7 The FAST and Stimulus Relatedness

As mentioned, the FAST is arguably a viable alternative to standard MTS testing procedures as a means of quantifying stimulus relatedness. The FAST's stimulus presentation and response formats differ largely from the MTS, and would therefore avoid the issues associated with the MTS for testing experimentally-varied inter-class stimulus relatedness. In this sense, the FAST is amenable to the quantification of differential relatedness along both inter- and intra-class parameters. As well as this, the FAST's metric amalgamates response time and accuracy in a conceptually-coherent fashion, further suggesting its suitability as a means of quantifying stimulus relatedness. Conceptually, therefore, the FAST is potentially an ideal candidate to fill the relatedness-quantification lacuna within the field of stimulus equivalence. Up to this point, discussion of the FAST and other implicit measures has centred on the fact that one block is consistent, and the other inconsistent, with an assumed repertoire of the subject. However, interestingly, there is comparably little experimental research in relation to any “implicit” test which has controlled the learning history of subjects prior to testing. Therefore, demonstrating experimentally that FAST scores vary in accordance with stimulus relatedness would also provide insights relating to the precise nature of the phenomenon measured by implicit measures generally. The remainder of the current chapter will discuss the limited number of attempts thus far to assess laboratory controlled stimulus relations using implicit measures. The experiments reported in the current thesis will then be outlined.

1.7.1 Measuring Laboratory-Controlled Stimulus Relations

As mentioned, almost no published research has examined the efficacy of implicit tests in detecting laboratory-created and controlled stimulus relations. The exceptions to this statement are two studies examining the IAT procedure using laboratory-controlled stimulus relations, both of which showed effects suggesting IAT effects could be brought under the control of trained stimulus relations (Gavin, et al., 2008; Ridgeway et al., 2010; see also Hall, Mitchell, Graham & Lavis, 2003, for a relevant study tapping into the same behavioral process harnessed by most implicit measures). Gavin et al. (2008) trained 2 arbitrary nonsense syllables (either in blue or in red) to be associated with either sexual or aversive pictorial stimuli. Subjects were then trained to form two 3-member stimulus equivalence relations, each consisting of either the blue or red stimulus, and 2 other nonsense syllables. The equivalent stimuli were then tested in a behaviour-analytically-modified version of the Implicit Association Test. On one block of the test, subjects were required to respond with one key press for stimuli from the ‘blue’ equivalence class and sexual images, and another key press for stimuli from the ‘red’ equivalence class and aversive images (i.e., responding consistently with the trained relations). The second block of the test required red-sexual pairings and blue-aversive pairings (i.e., inconsistent with the trained relations). Subjects were more fluent in responding on the block which was in accordance with the stimulus relations that had previously been trained, relative to the block which was incongruous to their histories. As such, this study suggested that the establishment of competing relational contingencies between stimulus relations could be readily detected on an implicit-style test, and hence provided groundwork for the generation of a function-analytic, well-understood implicit psychological measure.

Ridgeway et al. (2010) represented an extension of the paradigm employed by Gavin et al. (2008). Specifically, Ridgeway et al. trained two 3-member stimulus equivalence classes consisting of arbitrary nonsense syllables. Subjects were then exposed to an association phase where one member of each equivalence class was printed in either blue or red, and associated with pictures of either plants or animals. Like Gavin et al., subjects then completed a behaviour-analytic version of the IAT. Extending from this, subjects also completed a stimulus equivalence training procedure following this first IAT run which was different from the formerly trained relations, and tested again with a follow-up IAT. Similar to Gavin et al., responses in the first IAT from Ridgeway et al. were more fluent when response requirements were congruent to the trained stimulus relations. As well as this, after the training of alternative stimulus relations, performances on the follow-up IAT were reversed, and in line with the newly-trained stimulus relations. As such, this study replicated the findings of Gavin et al. and incorporated a reversal element of the procedure. The extensive behavioural control exerted in Ridgeway et al. thus provided further support for a function-analytic account of effects on implicit measures.

The evidence provided by the two aforementioned studies laid the ground-work for the development of the Function Acquisition Speed Test. To date, two studies have examined the FAST procedure using stimuli under experimentally-controlled relatedness. O’Reilly et al. (2012) initially trained conditional discriminative stimulus relations using the MTS procedure for two stimulus pairs (i.e., A1-B1 and A2-B2). Using A1, B1, and two novel stimuli (N1 and N2), it was demonstrated that subjects found it easier to form functional responses with A and B sharing the same response and N1-N2 sharing another response than to form functional responses where A and N1 shared a response and B and N2 shared a response. O'Reilly et al. (2013) extended the O'Reilly et al. (2012) paradigm to assess stimulus equivalence relations (rather than simple discriminative relations). In this study, two 3-member equivalence classes were established using one-to-many MTS training (i.e., training A1-B1, A1-C1, A2-B2, A2-C2, establishing classes A1-B1-C1, A2-B2-C2, with equivalence relations between B1-C1 and B2-C2). Subjects then completed a FAST which contained the stimuli from the trained equivalence classes. One block of the FAST involved B1 and C1 sharing a response option, and B2 and C2 sharing a different response option. The second block involved orthogonal requirements (i.e., B1 and C2 sharing a response option, and B2 and C1 sharing a response option). As expected, subjects who formed the aforementioned stimulus equivalence relations found it easier to respond on the FAST when B1-C1 shared a response and B2-C2 shared a response (i.e., consistent with their learning history), and more difficult to respond when B1-C2 shared a response and B2-C1 shared a response (i.e., inconsistent with their learning history). These studies provided the first formulation of the FAST procedure, as well as the first attempts using the FAST to follow-up on the findings of Gavin et al. (2008) and Ridgeway et al. (2010) in formulating a function-analytic account of effects on implicit measures.

In spite of the aforementioned work, in the entirety of the published literature there is no study to date which has systematically varied the relatedness of stimuli across experimental conditions and measured subsequent effects on implicit measures. Experimentally-addressing this question is salient, as it pertains to a highly relevant issue in the behaviour-analytic study of implicit measures. Specifically, it is unknown whether test effects noted in implicit measures are related in any way to the “strength” (i.e., relatedness) of the purported bias being measured. For example, if an IAT D-score of X is recorded, does this indicate stimulus relatedness (e.g., vegetables-good) that is half the level of that indicated by a D-score of 2X? This is a highly pertinent question, and social-cognitive researchers have been relatively conservative in discussing the meaning of IAT test scores (Greenwald & Nosek, 2008). While studies have sought to correlate implicit test scores with scores on other independent explicit tests for the purposes of convergent validity (see Teige-Mocigemba, Klauer, & Sherman, 2010, for a discussion), such studies do not provide a particularly reliable answer to this question, because the variables against which they correlate implicit test scores are themselves variable and uncontrolled. What is needed is an assessment of the variance in test effects as a function of experimentally-controlled or known (rather than inferred, as is the case in typical ‘known-groups’ approaches) degrees of stimulus relatedness. A first important step in understanding the behavioral processes involved in an implicit-style test for stimulus relatedness is to systematically manipulate the relatedness of laboratory controlled stimulus pairs, and observe the effects of these manipulations on test results. As such, the experiments outlined in the current thesis are not only of pertinence to the literature on stimulus relatedness quantification in stimulus equivalence research, but also in answering a pressing question in the literature on implicit testing.

1.7.2 The Current Research

The current research sought to assess the suitability of the FAST as a measure of stimulus relatedness in terms of both inter- and intra-class differences. Experiment 1 attempted to provide the first demonstration of the FAST’s ability to quantify stimulus relatedness varied along inter-class parameters. This involved assessing whether effects on the FAST varied systematically based on the relatedness between stimuli across 4 conditions. The first condition involved entirely untrained word associate pairs. Conditions 2 and 3 involved stimuli trained in equivalence relations using the MTS procedure, iterated either once (condition 2) or three times (condition 3). Condition 4 involved the administration of a FAST using real word stimuli from natural language categories of known relational strength (i.e., based on free association norms). It was expected that the FAST procedure should produce stimulus relatedness indices that increased in tandem with the amount of stimulus equivalence training and testing provided. As well as this, it was expected that those stimulus relations taken from the vernacular would produce the largest stimulus relatedness indices using the FAST method.

Experiment 2 involved assessing the effectiveness of the FAST in detecting differential stimulus relatedness varied along intra-class parameters. Subjects were trained to form relational classes between stimuli, with classes varying in size (either 0, 1, or 2 nodes in distance between tail-end stimuli). Each subject completed a FAST following training. In condition 1, the FAST stimuli used were A-B stimuli (i.e., 0 nodal distance), condition 2 involved FAST A-C FAST stimuli (i.e., 1 nodal distance), and condition 3 used FAST A-D stimuli (i.e., 2 nodal distance). It was expected that FAST scores would decrease incrementally as nodal distance between the stimuli increased (i.e., a decrease in stimulus relatedness).

In keeping with the trend of continual development of the FAST discussed in 1.6.3, a secondary aspect of the current research sought to compare the standard method of quantifying the fluency of responses in the FAST (i.e., through differences in the slope of the learning curve of each block) with a new method developed by the author which is based on a Skinnerian interpretation of the notion of ‘fluency’ (Skinner, 1937), as well as elements of assessment used within the precision teaching literature from applied behaviour analysis (Binder, 1996; Calkin, 2005). In brief, this method involves quantifying two rates of responding in each block: the rate of correct responding, and rate of incorrect responding.

The rate of incorrect responding is subtracted from the rate of correct responding, and this provides a response rate differential (RRD; or fluency) for the block. The response rate differential for the inconsistent block is subtracted from the response rate differential of the consistent block, and the difference in this value denotes the difference in the fluency of responding across blocks. This measure arguably offers even greater conceptual coherence to a behaviour-analytic interpretation of these effects than the current method of FAST fluency quantification, in that the response rate differential method avoids the use of linear modeling and the degree of mathematical abstraction notable in the latter.

The demonstration of changes in FAST fluency scores in accordance with both inter- and intra-class differences would provide evidence of the FAST's efficacy as means of quantifying stimulus relatedness, which would be conceptual and practical use for both stimulus equivalence researchers and implicit measures researchers. As well as this, the comparison of the difference in learning curves method with the aforementioned fluency differential approach will potentially provide the FAST with a novel, more conceptually coherent approach to quantifying fluency.

Chapter 2

Assessing the FAST as a Means of Detecting Differential Stimulus Relatedness along Inter-Class Parameters

Experiment 1

2. Experiment 1

2.1 Introduction

The current experiment involved the systematic manipulation of stimulus relatedness along inter-class parameters in order to assess the suitability of the FAST in quantifying such differences. The experiment involved four conditions. In the first condition, subjects received no MTS training before a FAST test using arbitrary stimuli. In the second and third conditions, subjects received one or three iterations of MTS training, respectively, before being presented with a FAST designed to assess the emergent relations. Finally, in the fourth condition subjects also received no MTS training but were exposed to a FAST using stimulus relations already established in the vernacular and of known association strength. It was intended that across the four conditions the varying levels of controlled stimulus relatedness would allow for a systematic examination of the variance in test effects as a function of relatedness.

It was expected that values for the four experimental DVs from the FAST (FAST score, rate-fluency differential score (RFD score; see Method), accuracy differences, and response times differences) would change linearly commensurate to the degree of relatedness between experimental stimuli, with greater relatedness leading to greater effect sizes. Further, it was conjectured that scores for block-slope and response rate differentials (RRD; see Method and 1.7.2) would change in accordance with an interaction of FAST block and training group. This interaction was expected due to the fact that increases in the strength of trained relations (i.e., an increase in positive stimulus control) ought to also lead to a decrease in the ability to parse those classes (due to increasing negative stimulus control). Thus, it was expected that scores on consistent FAST blocks would increase linearly as a function of training, while scores on the inconsistent FAST blocks would likely decrease in a similar fashion. Demonstration of the aforementioned findings would provide evidence for the FAST as a means of assessing inter-class differences in stimulus relatedness.

2.2 Method

2.2.1 Subjects

The subjects within this study (*n* = 62) consisted of Caucasian, Irish undergraduate students primarily attending Maynooth University, as well as three other nearby Universities. All subjects had normal or corrected-to-normal vision, and were not afflicted by any condition that may impair performance in tasks requiring sustained attention or learning. Subjects were allocated to one of four groups (‘No Training’, *n* = 12; ‘1 MTS Iteration’, *n* = 13; ‘3 MTS Iterations’, *n* = 20; and ‘Real Words’, *n* = 17). Forty one subjects identified as female, while the remaining twenty one subjects identified as male. Thirteen subjects did not pass equivalence testing within four training and testing cycles (see Procedure), and hence were excluded from the study. The remaining 49 subjects had a mean age of 20.44 years (*SD* = 1.771 years), and consisted of thirty three females and sixteen males; No Training *n* = 12, 1 Iteration *n* = 10, 3 Iterations *n* = 10, Real Words *n* = 17.

2.2.2 Ethical Considerations

The current experiment received full approval from the Maynooth University Ethics Committee. The tasks used in the current experiment posed no risks to subjects in terms of physical or mental health, and any stimuli used were either nonsense words or non-valenced words from the English language. No information on an individual level was recorded other than age and gender, and none of the experimental task performances provided any diagnostic information. All analyses in the current experiment were conducted at a group level. As such, participants’ data was anonymised. For ethical reasons a 40 minute time limit was placed on all subjects, as will be discussed below.

2.2.3 Apparatus

The experimental procedure was administered in a small, quiet research/study room (5’ X 5’ approx.) in Maynooth University. Subjects who were required to complete the matching-to-sample procedure in more than one sitting performed each subsequent training and testing iteration in the same research room. All subjects engaged in all procedures on a 13” *Apple MacBook* with a screen resolution of 1024 x 768 pixels. The MTS training as delivered using software created for this research using the experiment generation software PsyScope X (Cohen, MacWhinney, Flatt, & Provost, 1993), while the FAST procedure was delivered via proprietary software produced for this research using *Livecode.* All responses consisted of keyboard button presses or mouse-clicks, and all responses and their timings were recorded by the software programs.

Real word stimuli were chosen using the South Florida Free Association Norms Index (SFFANI; Nelson, McEvoy, & Schreiber, 1998). This index quantifies ‘association norms’ between stimuli based on discrete free association to a given exemplar stimulus. That is, the index lists correlation coefficients for word stimuli discriminating the response of any other given word stimulus. For instance, the stimulus ‘cheddar’ has a correlation coefficient of .922 with ‘cheese’ indicated a very strong correspondence between the vocal utterance ‘cheese’ and the presentation of the word ‘cheddar’. Interestingly, the presentation of ‘cheese’ is much less likely to produce the verbal response ‘cheddar’ and so the association value in this direction is represented by an r of .055. Four strongly-associated word pairs were selected for this study on the basis of having a normed association strength of at least $r = .3$ in *both* directions (this is relatively rare in the English language). The four selected pairs were pepper-salt (*r* = .695, .701), king-queen (*r* = .772, .73), washer-dryer (*r* = .755, .428), and sand-beach (*r* = .717, .394), where first and second *r* values in parentheses indicate forward and reverse association values, respectively.

2.2.4 Procedure

2.2.4.1 General Sequence

Subjects were quasi-randomly allocated to one of the four experimental conditions prior to commencement of experimentation. At the commencement of each experimental condition, subjects were asked to sign a consent form (which varied slightly depending on the condition to which subjects were allocated; see Appendix A and Appendix B) and were informed that they were free to withdraw from the experiment at any time. The first and fourth groups were required to complete a Function Acquisition Speed Test (FAST) with no training administered. The first group's FAST involved nonsense stimuli, with which subjects had no prior history. By contrast, the FAST completed by subjects in the fourth group used real word stimuli of strong relatedness (e.g. 'salt-pepper'). Two further groups received a specific number of iterations of a matching-to-sample training and testing procedure, which was designed to lead to the emergence of two stimulus equivalence relations between nonsense syllables. If MTS subjects did not reach criterion on both training and testing in any one session within 40 minutes total, the experimenter informed the subject that their participation was complete, debriefed them, and thanked them for taking part. This was the case for thirteen subjects. Following completion of the MTS procedure, subjects in the one-session group (i.e., Group 2) were then required to complete a FAST. Subjects in the 3 session group (i.e., Group 3) were required to complete the MTS procedure twice more in the same sitting (with approximate 2 minutes intervals between commencement of each iteration of the procedure), and were then required to complete the FAST. For all subjects, the FAST took no more than six minutes to complete. Inter-class stimulus relatedness was varied systematically via this stratification of groups, in that subjects who completed a greater amount of training would have had a greater probability to respond in accordance with the trained relations

2.2.4.2 Stimulus Equivalence Training and Testing

The matching-to-sample procedure was designed to establish two three-member equivalence relations among nonsense word stimuli for Groups 2 and 3. The stimuli CUG, JOM, VEK, LER, MAU, ZID were randomly assigned to their roles as samples and comparisons across the four MTS tasks, constituting two conditional discriminations. The two predicted emergent classes will be referred to here using the alphanumerics A1, B1, C1 and A2, B2, C2. The following instructions were provided on-screen to subjects at the commencement of each iteration of the MTS:

> In a moment some words will appear on this screen. Your task is to look at the word at the top of the screen and choose one of the two words at the bottom of the screen by “clicking on it” using the computer mouse and cursor. During this stage the computer will provide you with feedback on your performance. You should try to get as many answers correct as possible. Later on the task will become more difficult and feedback will no longer be presented. You will then need to rely on what you have learned during THIS stage of the experiment, so please pay close attention. If you have any other questions please ask them now. When you are ready please click the mouse button to begin the Experiment.

On all trials sample stimuli appeared at the centre-top of the screen in emboldened size 48 Times font, with two comparison stimuli appearing in the lower left and right corners of the screen one second following the presentation of the sample. There was no time limit on responding. In the training phase, each response was followed by the presentation of feedback in the form of the words ‘correct’ or ‘wrong’ appearing on the computer screen for one second. The feedback was accompanied by a brief auditory stimulus; a ‘beep’ sound for correct responses, or a deeper ‘buzz’ sound for incorrect responses.

Four stimulus relations were trained during the MTS phase (A1-B1, B1-C1, A2-B2, B2-C2). Consequently, there were four trial types: A1-B1 (B2), B1-C1 (C2), A2-B2 (B1), and B2-C2 (C1), where responses which resulted in negative feedback are parenthesised. Each of the four trial types was presented in a quasi-random order, with each trial presented once in a cycle of four trials. In effect, no trial type was presented more than twice in succession. Each trial type was presented eight times in total per training block, leading to a total of 32 trials per block of training. Subjects were required to reach a criterion minimum of thirty-one out of thirty-two correct responses in a block in order to progress to the equivalence test. If subjects failed to reach criterion on four consecutive training blocks, then the experiment was ended, and the subject was thanked for their participation.

Once criterion was met within four blocks of training, the test phase began immediately by the presentation of on-screen instructions. In this phase, the emergence of A1-C1, C1-A1, A2-C2 and C2-A2 relations was tested. Four trial types in this phase were present: A1-C1 (C2), A2-C2 (C1), C1-A1 (A2), and C2-A2 (A1). Erroneous responses in this phase are parenthesised. Each of the four trial types was presented in a quasi-random order, with each trial presented once in a cycle of four trials; again, no trial type was presented more than twice in succession. Each trial type was presented eight times in each iteration of this phase, leading to each testing block consisting of 32 trials.

No feedback was presented to subjects at any stage of the test phase. The criterion for passing this phase was also thirty-one out of thirty-two correct responses on a single block. If the test block was not passed, the subject was recycled back to training to criterion, again up to a maximum of four blocks, until criterion was met. At that point they were returned to testing. This continued in training and testing cycles until the test had been administered up to a maximum of four times. Fifteen subjects failed to pass the test after four testing cycles; when this occurred, subjects they were thanked for their time, debriefed, and their data was omitted from the study.

2.2.4.3 Function Acquisition Speed Test

Following the final exposure to the MTS procedure for those who received training, subjects who passed the testing phase were required to complete the Function Acquisition Speed Test (FAST). The FAST involved the formation of functional responses (i.e., press the ‘z’ or ‘m’ key) for particular stimuli presented in sequence on the computer screen. The procedure consisted of two blocks of 50 stimulus presentation trials each. One block was designated the ‘consistent’ block, and the other was designated the ‘inconsistent’ block. For those subjects in Groups 1 through 3 inclusive, the consistent block involved the reinforcement of responses, in the presence of a given nonsense stimulus, which would be consistent with those trained and tested in the MTS procedure (i.e., A1 and C1 share a response, and A2 and C2 share a response; note that although Group 1 received no training, the stimulus compatibilities were identical to those for Groups 2 and 3 but no FAST effect was expected). The inconsistent block involved reinforcing inconsistent functional response classes; that is, A1 and C2 shared a response, and A2 and A1 shared a response. See Figure 1 for a schematic of the functions trained for the A and C stimuli in each of the two FAST blocks. For subjects in Group 4, the consistent block involved the reinforcement of responses congruous with those associations from the SFFNAI (e.g., salt/pepper shared a response, and king/queen shared a response). The inconsistent block involved the opposite response patterns being reinforced (e.g., salt/king shared a response and pepper/queen shared a response). At the beginning of both blocks for all FASTs, the following text was presented to subjects:

> In the following section, your task is to learn which button to press when a word appears on screen. IMPORTANT: During this phase you should press only the Z key or the M key. Please locate them on the keyboard now. This part of the experiment will continue until you have learned the task and can respond without error. To help you learn you will be provided with feedback telling you if you are right or wrong. If you have any questions please ask the researcher now. Press any key when you are ready to begin.

Following the first key press of the subject, the first intertrial interval (ITI) of 500ms was presented (i.e., a blank white screen). Following this, a stimulus appeared in the centre of the screen in size 32 point font, after which point the subject was required to response with the ‘z’ or ‘m’ key. There was a 3000ms time limit on responding, after which the corrective feedback for an erroneous response was presented. If a response was made within 3000ms the screen was immediately cleared, and feedback was presented as appropriate. The order of block presentation on each FAST was randomized based on a random number generator built into the software. Both the consistent and inconsistent blocks of the FAST consisted of 50 trials; four different stimuli were presented in blocks: A1, A2, C1, and C2 for Groups 1 through 3, and two real-word pairs for Group 4. These stimuli were presented in a quasi-random order, with each stimulus presented once in a given cycle of four trials. Each block of the FAST consisted of twelve and a half cycles, with two stimuli presented an extra time in order to complete the block of fifty trials.

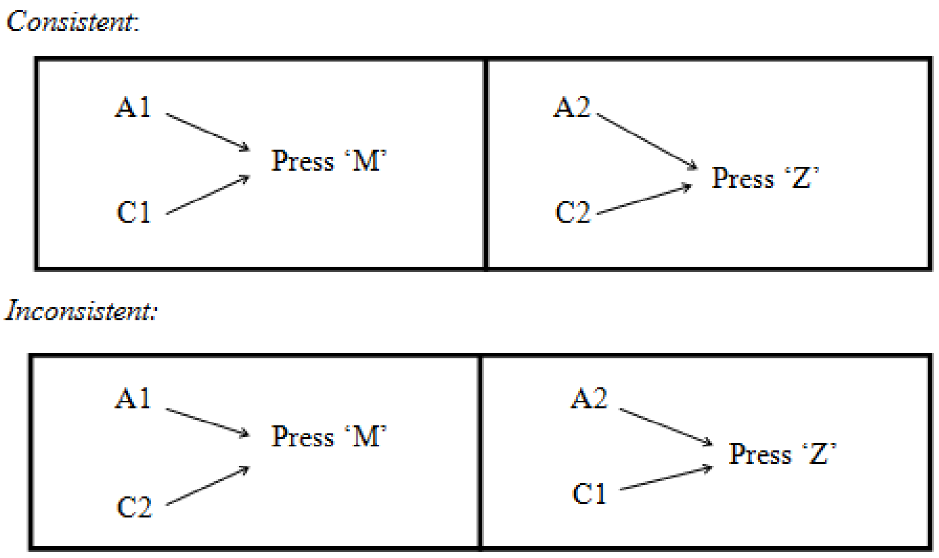


Figure 3 The trained response functions for FAST stimuli across both the consistent and inconsistent blocks.

2.3 Results

2.3.1 Measurement

Data for the MTS procedure were measured in terms of number of blocks of training and/or testing required to reach criterion. Fluency scores on the FAST were calculated using two metrics. The first was the standard FAST score. Linear regression lines were generated according to the cumulative record generated by responses on each block. The slope of the regression line recorded for the inconsistent block was subtracted from that recorded for the consistent block, and this value is henceforth referred to as the FAST score. A positive FAST score indicates a steeper learning rate on the consistent block compared to the inconsistent block, which would suggest that the functional classes established in the consistent block would be compatible with the stimulus equivalence training and testing contingencies. A negative FAST score indicates the opposite.

The second score was a novel measure of fluency developed by the author, as mentioned in 1.7.2. For each block, two rates of responding were calculated: the rate of correct responses per minute (CRPM), and the rate of incorrect responses per minute (IRPM). Readers should note that the choice of rate ‘per minute’ is arbitrary, but is in line with behaviour-analytic tradition. These rates were calculated by dividing the number of correct (for CRPM) or incorrect (for IRPM) responses by the total time the subject took to complete the FAST block[2], and multiplying this value by 60,000 (i.e., to translate the value to ‘per minute’). To provide a composite index of learning rate, the IRPM rate was then subtracted from the CRPM rate, providing what can be called a response rate differential (RRD) score for the block. This RRD has face validity as a representation of the fluency of responding insofar as fluency is often described as a high rate of correct responding, and a low rate of incorrect responding (Binder, 1996). To produce a FAST score, the RRD of the inconsistent block was subtracted from the RRD of the consistent block, and this value represented the fluency difference for that subject’s FAST, herein referred to as the rate-fluency differential (RFD) score (see equation 1 below for an alternative expression of this formula, where TC = Total Correct, TI = Total Incorrect, TT = Total Time, and where a ‘C’ subscript indicates the consistent block, and an ‘I’ subscript indicates the inconsistent block). A positive RFD score indicates that subjects were more fluent on the consistent block, while a negative RFD score indicates that subjects were more fluent on the inconsistent block.

$$RFD = \left(\left(\frac{TC_C - TI_C}{TT_C}\right) - \left(\frac{TC_I - TI_I}{TT_I}\right)\right) \times 60000 \qquad (1)$$

Equation 1. Formula to calculate the rate-fluency differential score.

---

[2] It should be regarded that both intertrial interval and feedback presentation times are included within the total time taken to complete the procedure. The rationale for their inclusion is based on the fact that the measure is attempting to describe the true rate of the behaviour *within the procedure*. Latent periods (such as ITIs and feedback presentation) where the subject cannot respond necessarily constrain the overall rate of responding in a given block, and so are included.

2.3.2 Descriptive Statistics

2.3.2.1 Matching-to-Sample

Initial descriptive analyses showed that as subjects received more iterations of the MTS procedure, they tended to need less training and testing block cycles in order to complete the procedure. See Appendix C for full illustration of MTS data for each subject, as well as the number of training and testing phases required to complete the procedure at each iteration that the subjects completed.

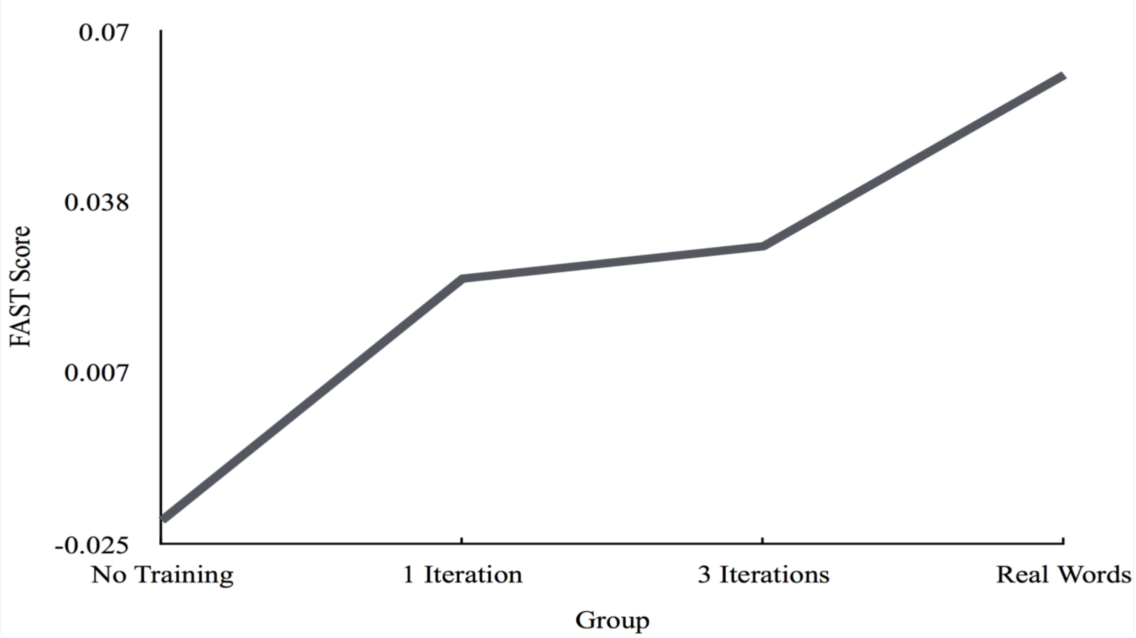
2.3.2.2 Function Acquisition Speed Test

Descriptive analyses showed further that, as expected, more stimulus equivalence training iterations was associated with larger FAST scores and larger RFD scores. Table 3 displays the means for FAST scores, RFD scores, as well as raw accuracy and response time inter-block differences for each of the four groups. Figure 2a shows the trend of the FAST scores across groups, while Figure 2b shows the trend of RFD scores across groups. It can be seen that with increasing relatedness, larger FAST scores and RFD scores are apparent.

Table 3. Table illustrating the mean FAST score, RFD score, accuracy differences, and response time differences (in milliseconds) recorded by participants from different training conditions on the FAST. A positive score indicates a faster/more accurate/more fluent consistent block, while a negative score indicates the opposite. For instance, a mean difference of –93.41 on response time for the no training group indicates that subjects were on average 93.41 milliseconds faster on the inconsistent block compared to the consistent block. The numbers in parentheses indicate standard deviations.

	Fast Score	RFD Score	Accuracy Difference	Response Time Difference
No Training	-.02 (.06)	-1.7 (4.35)	-1.33 (4.38)	-93.41 (208.05)
1 Iteration	.024 (.07)	1.48 (6.19)	1.1 (5.55)	-89.95 (185.67)
3 Iterations	.03 (.063)	2.3 (4.19)	2.3 (3.74)	-9 (199.46)
Real Words	.06 (.1)	6.37 (9.28)	4.82 (8.77)	207.66 (260.92)

4a.



4b.

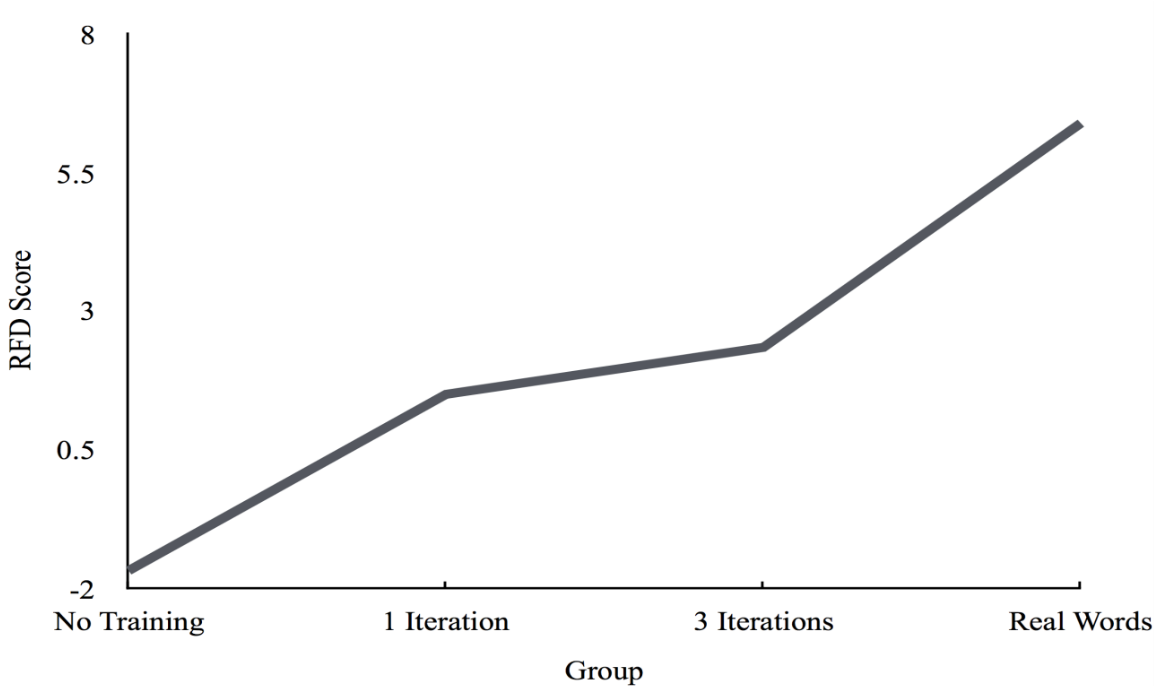


Figure 4 Figure 4a shows the trend of FAST scores as a function of training group across the four conditions, while Figure 4b shows the trend of RFD scores across the same conditions.

Further descriptive analyses of performances within the FAST blocks revealed what appeared to be a general improvement in response time, accuracy, FAST scores, and RFD scores on *both* consistent and inconsistent blocks across conditions. As Table 4 illustrates, on each block, mean FAST scores, RRD (i.e., response rate differential, the difference between CRPM and IRPM on an *individual* block) scores, and accuracy scores in general tended to be higher, and response times lower as purported relatedness increased.

Table 4. Table showing the different mean scores on Block-Slope Score, Response Rate Differential, Accuracy, and Response Time in both the consistent and inconsistent blocks recorded for subjects in each condition. Response time is in ms, and all parenthesized values are standard deviations.

		Consistent	Inconsistent
Block-Slope Score	No Training	.323 (.09)	.343 (.08)
	1 Iteration	.412 (.03)	.388 (.07)
	3 Iterations	.433 (.03)	.403 (.06)
	Real Words	.419 (.05)	.357 (.09)
RRD Score	No Training	13.29 (7.88)	14.99 (6.98)
	1 Iteration	19.77 (3.74)	18.29 (6.16)
	3 Iterations	21.85 (3.03)	19.52 (4.77)
	Real Words	21.76 (3.73)	15.39 (8.3)
Accuracy	No Training	38.5 (7.19)	39.8 (6.09)
	1 Iteration	43.7 (3.37)	42.6 (5.36)
	3 Iterations	45.6 (2.07)	43.3 (3.8)
	Real Words	44.76 (3.23)	39.94 (7.77)
Response Time	No Training	1038.01 (323.4)	944.61 (232.58)
	1 Iteration	773.58 (101.19)	863.53 (206.85)
	3 Iterations	780.89 (218.25)	771.89 (116.41)
	Real Words	685.28 (130.69)	892.94 (249.74)

2.3.3 Inferential Statistics

2.3.3.1 Frequentist Analyses

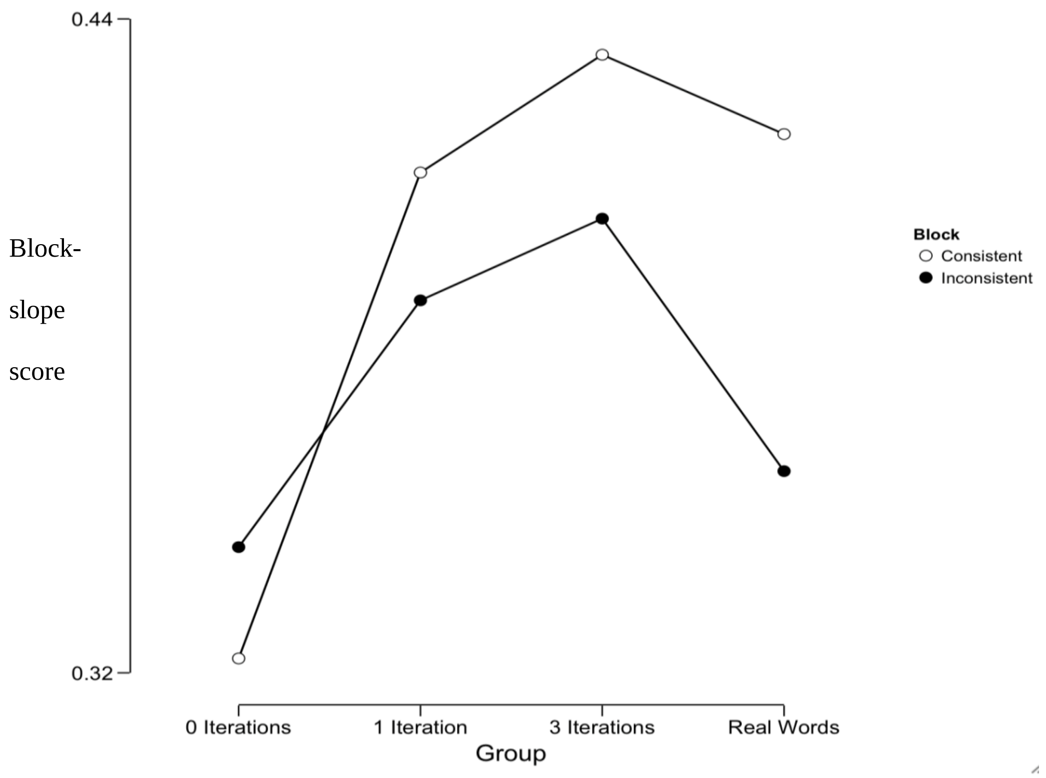
Inferential analyses were run to determine whether there was any significant effect of condition on FAST scores, RFD scores, block accuracy, or mean block response time. The former three measures were not normally distributed. Response times were normally distributed, however. Mean response times also did not violate the assumption of homogeneity of variance (p = .704). Given that it was expected that effects would increase as a function of stimulus relatedness, linear polynomial contrasts were run using each of the four metrics. In essence, a linear polynomial contrast measures how well specific data points fit a linear trend. A significant linear trend as a function of group was found for FAST score (F (1, 45) = 11.516, p = .00145; η_p^2 = 0.204, observed power = .913), RFD score (F (1, 45) = 13.366, p = .0007; η_p^2 = 0.229, observed power = .947), accuracy (F (1, 45) = 9.644, p = .0033; η_p^2 = 0.176, observed power = .860), and response times (F (1, 45) = 9.451, p = .004; η_p^2 = 0.174, observed power = .853).

Two analyses of variance were run in order to further probe the nature of the relationship between condition and block on fluency scores, one using the block-slope score, and the other using the RRD score. That is, two mixed 2x4 within-between subjects ANOVA was run in order to further analyse these effects. The reader should bear in mind when interpreting these findings that the assumption of normality was not maintained for either measure of fluency. The assumption of homogeneity of intercorrelations according to Box’s M was maintained for both FAST score (F (9, 13045.057) = 2.286, p = .015, α = .001) and RFD scores (F (9, 13045.057) = 2.62, p = .005, α = .001). Contrary to what was expected, for FAST score, no significant interaction effect was found between FAST block and training condition, Wilks’ Lambda = .852, F (1, 45) = 2.610, p = .063, η_p^2 = .148, observed power =

.602. A significant, medium main effect was found for FAST block, Wilks' Lambda = .913, *F* (1, 45) = 4.311, *p* = .044, η_p^2 = .087, observed power = .529. As well as this, a strong, significant main effect was found for Group, *F* (3, 45) = 4.961, *p* = .005, η_p^2 = .249, observed power = .887 (see Figure 3a). When RFD score was used as DV, a significant interaction effect was found between FAST block and training condition, Wilks' Lambda = .814, *F* (3, 45) = 3.428, *p* = .025, η_p^2 = .186, observed power = .734 (see Figure 3b).

A final frequentist analysis was run in order to determine whether there were any order effects involved in FAST block presentation, the presence of which may confound the veracity of the current data. As such, four Mann-Whitney U Tests were run, each of which used FAST block order as IV, and Fast score, RFD score, accuracy, and response time as DVs, respectively. No significant order effects were found for any four mentioned DVs (*p*s = .854, .927, .936, and .713, respectively).

5a.



5b.

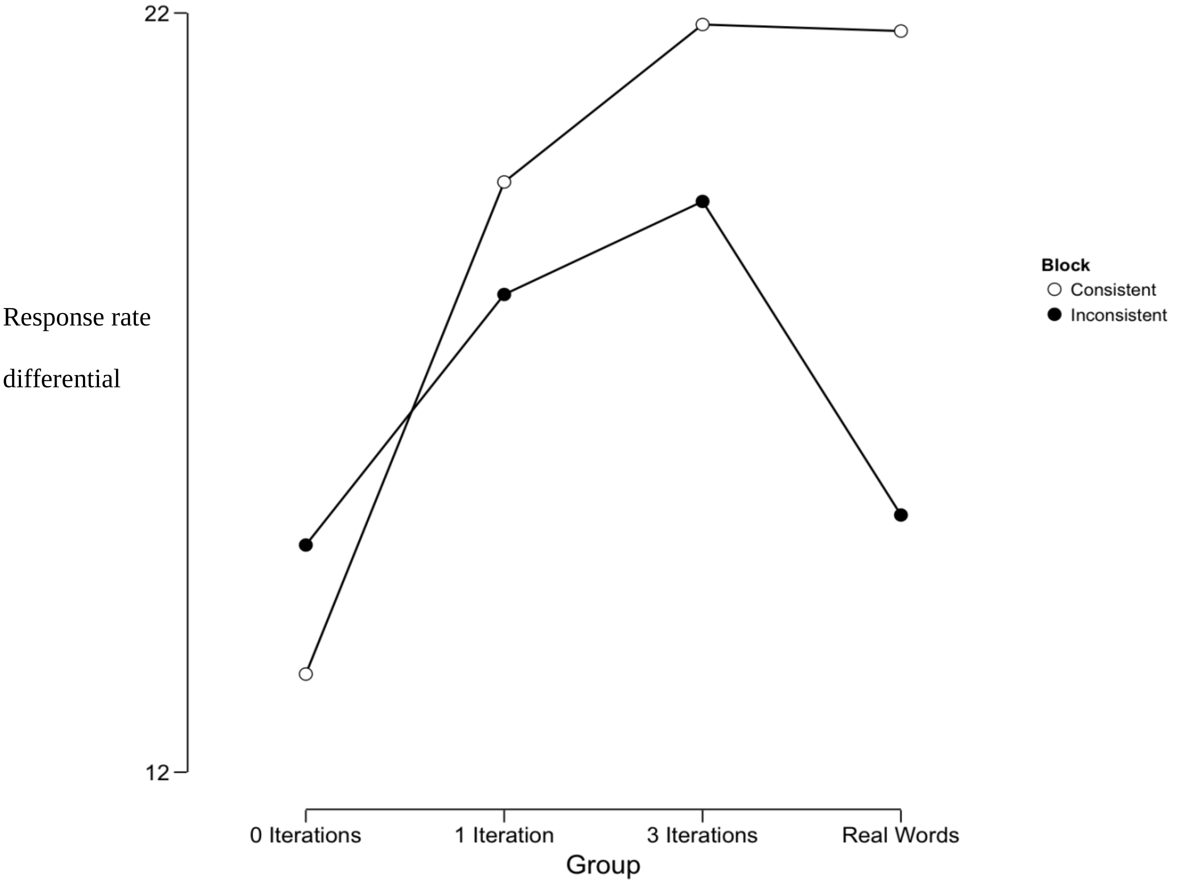


Figure 5. Figure 5a outlines the trend of both consistent and inconsistent block-slope scores as a function of experimental condition. Figure 5b illustrates the trend of response rate differentials for the consistent and inconsistent blocks across experimental conditions.

2.3.3.2 Bayesian Analyses

Bayesian inferential statistics were also run in order to provide a more detailed overview of the data. The Bayesian approach to statistical analysis is conceptually distinct to the Frequentist approach. Specifically, Frequentist analyses may be understood as investigating the probability of the data given a hypothesis (e.g., there is an effect), while Bayesian analyses investigate the probability of a hypothesis given the data (Gelman, et al., 2014). While both conceptual approaches have merit, it may be argued that the Bayesian approach is more amenable to an inductive scientific philosophy (as that of experimental behaviour analysis), in that Bayesian analyses allow the incorporation of prior observations, as well as the current, in assessing the veracity of a particular conjecture. This argument aside, supplementing Frequentist analyses with Bayesian analyses can provide a more comprehensive picture of the analysed data, in particular the effect sizes of findings (Valentine, Buchanan, Scofield & Beauchamp, 2017).

The main output of Bayesian analyses is a Bayes Factor[3], which simply refers to the relative evidence based on the given data for one model (i.e., an effect exists), compared to another (i.e., an effect does not exist). Four Bayesian one-way ANOVAs were run in order to assess whether models assuming (respectively) that FAST score, RFD score, accuracy, or response time varied as a function of relatedness were more likely than the null hypothesis. For FAST score, $BF_{10} = 1.372$, indicating that these data are more 1.372 times more likely to be observed given the existence of an effect than given the lack of existence of an effect (all further Bayes Factors may be interpreted in a similar fashion). Based on the criteria of Lee and Wagenmakers (2013), this represents ‘anecdotal’ evidence for the likelihood of an effect. For RFD score, $BF_{10} = 3.028$, indicating moderate evidential value. For accuracy, $BF_{10} = 0.995$, indicating no evidence. Finally, for response times, $BF_{10} = 9.954$, indicating moderate-to-strong evidence.

As well as the former analyses, two 2x4 mixed between-within groups Bayesian ANOVAs were run, each using block as within subjects factor, and group as the between subjects factor. One ANOVA used block-slope score as DV, while the other used RRD score. For the ANOVA with block-slope score as DV, there was anecdotal evidence for an interaction effect between block and group ($BF_{10} = 1.628$). For the ANOVA with RRD score

---

[3] Though beyond the scope of the current thesis, it should be noted that the use of Bayes factors has been a point of contention in modern statistical theory. For detailed discussion on this point, see Morey, Romeijn, and Rouder (2016).

as DV, there was moderate evidence for an interaction effect between block and group ($BF_{10} = 3.64$).

2.3.4 Summary

A general increase in all four relevant metrics (FAST score, RFD score, accuracy difference, and response time difference) was seen with greater relatedness between procedural stimuli. These increases were significant based on the inferential analyses, with RFD score yielding the largest frequentist effect size. Bayesian analyses indicated that only the models from RFD score and response time provided evidence above an anecdotal level for an effect. Unexpectedly, FAST block-slope scores did not change significantly as a function of interaction between FAST block and training condition; however, main effects were noted for both block and training condition, and the power achieved by the analysis was relatively low. RFD scores changed as a function of interaction between FAST block and training condition, with consistent block scores generally increasing, and inconsistent block scores exhibiting a quadratic trend (i.e., increasing, then decreasing). The Bayesian models for these analyses confirmed the above, indicating that RRD score’s interaction was above an anecdotal level of evidence, while block-slope score’s was not. None of the above analyses were confounded by any FAST block order effects.

2.4 Discussion

The current experiment intended to assess whether stimulus relatedness varied along inter-class parameters would lead to commensurate changes in scores in the Function Acquisition Speed Test. Relatedness was varied across 4 levels: No Training, 1 MTS Iteration, 3 MTS Iterations, or Real Word stimuli. It was expected that FAST scores, RFD scores, mean accuracy block-differences, and mean response time block-differences would increase as a function of increasing stimulus relatedness. It was also expected that FAST block and training condition would interact in informing block-slope scores, and RRD scores.

The sets of both Bayesian and Frequentist analyses showed that scores on all four metrics changed in the expected direction commensurate with training condition. The largest effect noted in the Frequentist analyses was for RFD score, while the largest effect noted in the Bayesian analyses was for response time differences. However, based on the combination of the evidence provided by both Bayesian and Frequentist analyses, RFD scores proved to be the most sensitive measure to changes in stimulus relatedness in the FAST. Further, an interaction between FAST block and training condition was noted when subjects' FASTs were scored using response rate differentials on each block, achieving significance in the Frequentist analyses, and suggesting moderate evidence in the Bayesian analyses. No such interaction was found for block-slope scores in either the Frequentist or Bayesian analyses (though two main effects were found in the Frequentist analyses for this metric). Notable also is that throughout the Frequentist analyses, RFD score consistently achieved greater statistical power than FAST score (likely as a function of its larger effect size).

The current experiment demonstrated that scores in the FAST change commensurate to the degree of experimentally-varied relatedness along inter-class parameters between test stimuli. Interestingly, the rate-fluency differential metric proved to be more sensitive than the traditional FAST score measure in detecting these changes. A possible explanation for the relative effectiveness of RFD to FAST score may be understood in terms of the relative weighting of the two compound scores (i.e., response time and accuracy) in both measures. Firstly, note that in terms of the cumulative evidentiary value of both Bayesian and Frequentist analyses, response time was more greatly influenced than accuracy by the degree of relatedness between experimental stimuli (i.e., both were significant and showed near-identical effect sizes in the Frequentist analyses, but response time's BF was nearly 10, while accuracy’s was essentially 1). Note further that this response time measure is based on mean response times (mean RT), rather than mean of cumulative time taken (mean CTT) in the procedure as is taken in FAST score/RFD score.

In the analysis of response time, substituting mean RT for mean CTT would have no impact on statistical findings, given that the absolute difference of these scores would remain the same. That is, mean CTT essentially involves the addition of some fixed temporal additive to each trial’s time value (i.e., 1500ms, including intertrial interval and feedback presentation). As such, in the current design, mean CTT will always equal mean RT plus 1500ms. Thus, the absolute difference between mean CTT scores for blocks will be the same as the difference between mean RT scores for those same blocks (given that the addition of 1500ms to both blocks essentially cancels the other out). By contrast, the absolute difference in RRD scores (i.e., RFD score) does not remain the same when the temporal additive to mean RT changes. Specifically, the CTT score is taken as the denominator of the equation in the formula used to calculate response rate differentials for blocks. As such, increases in the value of the denominator across RRD scores *do* result in absolute differences in RFD scores. For example, the difference between 1/2 and 1/3 (i.e., 0.1666) is not the same as the difference between 1/3 and 1/4 (i.e., 0.08333), in spite of the absolute difference in the size of the denominator being the same (i.e., 1). Note further that, all other factors being equal, a change in the value of the denominator become less impactful on the absolute difference in scores as the denominator becomes larger (i.e., the function $f(x) = 1/x$ is asymptotic). Given that mean CTT involves the addition of a fixed time onto each block, it may be suggested from this that the effect of differences in response times in this measure is diminished relative to the mean RT measure.

Consider further the FAST score. The FAST score is calculated through the difference in block-slope scores of each FAST block. The slope (m) of the regression line is calculated as:

$$\mathrm{m} = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \qquad (2)$$

Equation 2. Formula for the calculation of the slope of a regression line fitted to a series of data points.

where N refers to the number of observations (i.e., trials), X refers to plots on the X axis (i.e., cumulative time), and Y refers to plots on the Y axis (i.e., accuracy). Explicating on the specific nuances of the dynamical changes which occur within this formula at each step of increasing values of X (and the relative interaction of X and Y changes) is beyond the scope of the current thesis. However, in brief, the presence of the X values in both the numerator and denominator of the equation saliently influences the relative effect of changes in response times. That is, in a similar fashion to that of RFD score, increasing overall response times (via inclusion of intertrial intervals and feedback) reduces the effect of differing cumulative response times. However, with FAST score, this is further compounded by the use of the product of X values with Y values in the numerator. As such, the impact of differing response times is even less salient than that of RFD, which arguably results in the diminution of effects seen relative to RFD. The veracity of this position can, and will, be assessed in Experiment 2. Specifically, it would be expected that if response times are influenced to a greater extent than accuracy, then RFD score should prove to be more effective in measuring differences. Conversely, if accuracy is more strongly affected than response time, then mean FAST scores should exhibit a greater effect size than that of RFD score. Nevertheless, this is largely a psychometric question, and as such, future studies should seek to assess the psychometric properties of these metrics, and the specific dynamics at play in their production.

Of interest in the current findings is the trend of scores on both the consistent and inconsistent blocks across different degrees of stimulus relatedness. It is notable that while the consistent block shows a stable score between the 3-Iteration and Real Word conditions, the inconsistent block shows a decline in fluency from the 3-Iteration condition to the Real Word condition. As such, it may be suggested that the general increasing difference between block scores is primarily the consequence of this relative decline in inconsistent scores, rather than an increase in consistent scores. This is in line with previous research findings (Cartwright, 2013), in that the FAST appears to more readily quantify resistance to behavioural change, rather than behavioural facilitation. A possible explanation for this may be due to a ‘ceiling effect’ of procedural simplicity. Specifically, it may be the case that the FAST procedure is relatively simple for subjects to complete, and so fluency is quickly achieved after very little training in the relevant stimulus relations. Notably, the fluency of both blocks initially increases (albeit to a greater degree for the consistent block) from No Training to 1-Iteration. In terms of a stimulus control topography, it may be suggested that at this early stage of learning, the stimuli are under some degree of positive stimulus control (i.e., S+ control), but little-to-no negative stimulus control (i.e., S- control; McIlvane & Dube, 1992). As such, the formation of functional responses for stimuli within the FAST is generally facilitated by only a slight change in the degree of relatedness between experimental stimuli. However, as the degree of relatedness between the experimental stimuli continues to increase, it becomes more difficult for subjects to pair stimuli orthogonally to the trained/derived relations. This would suggest that greater degrees of stimulus relatedness lead to an increase in S- control of the experimental stimuli (i.e., through increases in resistance to behavioural change), with the increase in differences between blocks in the FAST denoting this increase in S- control.

It is notable that S- stimulus control was not explicitly trained within the current procedure. While subjects were punished for erroneous responses (which may be argued as being some degree of S- control), the instantiation of more explicit S- control may lead to even greater effects than those noted in the current experiment. In particular, those ‘Real Word’ stimuli used in the current experiment were of great S+ control, but comparably little (if any) negative stimulus control (i.e., ‘salt’ is strongly related to ‘pepper’, but is not strongly opposed to ‘queen’; it is merely unrelated). A method of investigating this conjecture regarding the effect of explicit negative stimulus control training would be to introduce a procedure which is capable of actively reinforcing both S+ and S- control. This may be difficult to achieve, as attempting to reinforce responding away from a given stimulus may inadvertently instead reinforce S+ control with the other alternatively-selected stimulus. Alternative procedures to the MTS may be more viable as methods of training S- control, such as responding to the veracity of propositions based on stimulus relations (e.g., reinforcing the selection of ‘TRUE’ when presented with the proposition ‘A1 is not the same as C2’). For example, an adaption of the format of the Relational Evaluation Procedure (REP; Stewart, Barnes-Holmes, & Roche, 2004) may provide such a procedure. The use of this method (as an alternative to MTS) in the vein of the design of the current experiment would allow researchers to more actively control S-S relations, and would grant further insight into the precise outcomes of varying stimulus control topographies (and, by extension, stimulus relatedness) on FAST scores.

The results of the current experiment suggest that effects in the field of implicit measures are the result of the degree of relatedness of the procedurally-relevant stimuli in the learning history of the subject. In this sense, the current experiment provides the first empirical support for a long-held assumption of the implicit measures literature. However, a number of further issues remain unexplored. For example, it is unclear what effect further

MTS training would have on scores in the FAST. The aforementioned discussion of the potential effects of incorporating greater S- control may provide some degree of insight into this. However, it is notable that the ceiling effect for the consistent block was present after only 3-iterations of MTS training. That is, the scores for the consistent block for the 3-Iteration condition versus the Real Words condition was effectively the same, in spite of the Real Words condition having stimuli of far greater relatedness. As such, it may be the case that the consistent block of the FAST is relatively easy for subjects irrespective of the degree of relatedness between stimuli. It could be argued from this that stimuli which are widely different in terms of relatedness in the repertoire of subjects may show no differences on the consistent block of the FAST. In order to address this issue, it may be beneficial to adjust the relative difficulty of the FAST. For example, instantiating shorter response windows may increase the difficulty of both FAST blocks, leading to a greater degree of variance in scores, and thus more effectively capturing differences in S+ control (though see the Discussion chapter for an overview of the relationship between the relative effect of differing response windows and its relationship to time/accuracy). While the aforementioned issues (as well as other points) will be more thoroughly explicated in Chapter 4, the current experiment demonstrated that differential inter-class stimulus relatedness lead to commensurate differences in the scores on the Function Acquisition Speed Test.

Chapter 3

Assessing the FAST as a Means of Detecting Differential Stimulus Relatedness along Intra-Class Parameters

Experiment 2

3. Experiment 2

3.1 Introduction

The current experiment intended to investigate whether differential stimulus relatedness which was varied along intra-class parameters would lead to proportionate changes in scores on the FAST. A number of studies in the field of stimulus equivalence have noted that increasing nodal distance between stimuli in stimulus classes leads to a reduction in the degree of transfer of function between those stimuli, as well as a decrease in the probability of equivalence class formation (i.e., a decrease in relatedness; Fields et al., 1995; 2007; 2008). Further, Bentall et al. (1999) noted that increasing nodal distance between stimuli lead to both an increase in errors on testing, and an increase in response latencies for matching equivalent stimuli. If the FAST is to be purported as a measure of stimulus relatedness across both inter- and intra-class relatedness, then scores in the procedure should change in accordance with differing nodal distance. The nodal differentiation of probed stimulus relations across FASTs in the current experiment was identical to that of Bentall et al. (i.e., 0-, 1-, or 2-node classes). Unlike Experiment 1, whose design was (necessarily) based solely on a between-groups research question, the current experiment focused on within-subject changes in scores. That is, all subjects were trained with identical stimulus relations, and tested with three different FASTs involving stimuli of differing nodal distance. Given that the typical stimulus equivalence testing format used in MTS studies may actively participate in the formation of stimulus equivalence classes (Cummins, Roche, Tyndall, & Cartwright, in press), the current experiment opted to administer the standard MTS format of equivalence testing only after each of the FASTs was completed.

It was expected that increasing nodal distance between experimental stimuli in the FASTs would lead to a decrease in the difference between scores on the FAST blocks. As well as this, it was expected that FAST blocks and nodal distance would interact in informing the values of FAST block-slope scores and RRD scores. A comparison of the aforementioned two metrics was a secondary goal of the current experiment, given the relative efficiency of RFD relative to FAST score in the previous experiment. As discussed in 2.4, it could be expected that if accuracy is more greatly affected by differential intra-class stimulus relatedness than response time, then FAST score ought to be more sensitive than RFD score. Conversely, if response times are affected more greatly than accuracy, then it should be expected that RFD score will exhibit greater effect sizes than FAST score.

3.2 Method

3.2.1 Subjects

The subjects within this study (n = 16) consisted of Caucasian, Irish undergraduate students attending Maynooth University. Subjects were recruited through the use of the Maynooth University participant pool, and received no remuneration for their participation. All subjects had normal or corrected-to-normal vision, and were not afflicted by any condition that may impair performance in tasks requiring sustained attention or learning. All subjects received MTS training, three FASTs in varying orders (which individually probed for either A-B, A-C, or A-D stimulus relatedness), and then a stimulus equivalence testing procedure which tested for the aforementioned stimulus equivalence relations. Nine subjects identified as female, while the remaining seven subjects identified as male. The subjects had a mean age of 20.94 years (*SD* = 1.73 years).

3.2.2 Ethical Considerations

The current experiment received approval from the Maynooth University Ethics Committee. The tasks used in the current experiment posed risks to subjects in terms of physical or mental health, and any stimuli used were either nonsense words or non-valenced words from the English language. No information on an individual level was taken from subjects, and none of the experimental tasks provided any diagnostic information. All participant data was anonymised. It was possible that subjects may find the MTS procedure to be tedious; in order to prevent this, a limit on the number of cycles of each training block was placed on all subjects.

3.2.3 Apparatus

The experimental procedure was administered in a small, quiet research/study room (5’ X 5’ approx.) in Maynooth University. All subjects engaged in all procedures on a 13” *Apple MacBook* with a screen resolution of 1024 x 768 pixels. The MTS training and testing procedures were delivered using software created for this research using the experiment generation software PsyScope X (Cohen, MacWhinney, Flatt, & Provost, 1993), while the FAST procedure was delivered via proprietary software produced using *Livecode.* All responses consisted of keyboard button presses or mouse-clicks, and all responses and their timings were recorded by the software programs. All stimuli in the current experiment consisted of three-letter, mono-syllabic nonsense words.

3.2.4 Procedure

3.2.4.1 General Sequence

At the commencement of the experiment, subjects were asked to sign a consent form (see Appendix D) and were informed that they were free to withdraw from the experiment at any time. All subjects completed a match-to-sample procedure designed to lead to the emergence of two 4-member stimulus classes between nonsense syllables. If subjects did not reach criterion on the MTS simultaneous training phase, the experimenter informed the subject that their participation was complete, debriefed them, and thanked them for taking part. However, all subjects reached criterion in the simultaneous phase of training. Following completion of the MTS procedure, all subjects were then required to complete three FASTs. Each FAST used a pair of stimuli from each of the trained classes. The pairs of stimuli differed across the three FASTs in terms of the nodal distance between the stimuli (i.e., in the 0-node FAST, A-B relations were tested; in the 1-node FAST, A-C relations were tested; in the 2-node FAST, A-D relations were tested). For all subjects, each FAST took no more than six minutes to complete. Subjects then completed a MTS stimulus equivalence testing procedure, which consisted of 60 MTS trials which probed for relevant stimulus relations (i.e., A-B/B-A, A-C/C-A, A-D/D-A).

3.2.4.2 Stimulus Equivalence Training

The current experiment employed a linear matching-to-sample procedure which was designed to establish two 4-member equivalence classes for subjects. These relations were first trained sequentially. That is, A1-B1/A2-B2 relations were first trained to a criterion of 15/16 correct responses in a block (16 was chosen as this represented 4 cycles of each of the 2 trial configurations). Then, B1-C1/B2-C2 relations were trained to an identical criterion as the A-B condition. Following this, C1-D1/C2-D2 relations were also trained to the same criterion. When criterion for each of these stimulus relations was met, a simultaneous training block was administered. That is, a block presented all of the aforementioned trials in a single block, with each trial cycled three times per block. As such, subjects were required to achieve a criterion of 17/18 correct in this block (18 was chosen as this represented 3 cycles of each of the 6 relevant trial configurations). The predicted emergent classes were A1-B1-C1-D1 and A2-B2-C2-D2. The arbitrary stimulus exemplars used in the current experiment consisted of CUG, VEK, JOM, KAH, MAU, ZID, LER, and YOH. The following instructions were provided on-screen to subjects at the commencement of the MTS:

> ***PLEASE PAY ATTENTION WHEN READING THE FOLLOWING INFORMATION***
>
> In a moment some words will appear on this screen. Your task is to look at the word at the top of the screen and choose one of the two words at the bottom of the screen by clicking on it using the computer mouse and cursor. More stages will follow after this, with the same task using different words each time.
>
> During all stages the computer will provide you with feedback on your performance. Use this feedback to learn how to respond. You should try to get as many answers correct as possible.
>
> If you have any questions please ask them now.
>
> When you are ready please click the mouse button to begin the Experiment.

On all trials sample stimuli appeared at the centre-top of the screen in emboldened size 48 Times font, with two comparison stimuli appearing in the lower left and right corners of the screen one second following the presentation of the sample. There was no time limit on responding. In all blocks of the training procedure, each response was followed by the presentation of feedback in the form of the words ‘correct’ or ‘wrong’ appearing on the computer screen for one second. The feedback was accompanied by a brief auditory stimulus; a ‘beep’ sound for correct responses, or a deeper ‘buzz’ sound for incorrect responses.

As mentioned, the MTS training consisted of two main stages: the sequential phase, and the simultaneous phase. Two stimulus relations were trained during the first block of the sequential phase (A1-B1, A2-B2). In this block, the sample stimulus was always an A stimulus, and the comparison stimuli were always the B1 and B2 stimuli. Following the achievement of criterion, subjects were provided with a short break by the procedure, and upon their prompting (via a mouse click), moved on to the second block of training. This second block involved the training of two further stimulus relations (B1-C1, B2-C2). In this block, the sample stimulus was always a B stimulus, and the comparison stimuli were always the C1 and C2 stimuli. When the subject reached criterion here, they moved on to the third block, which involved identical configurations to the previous two blocks, but trained C1-D1 and C2-D2 stimulus relations. When criterion was reached after this third block, the subject moved on to the simultaneous phase of the training. This involved the further training of all six of the previous stimulus relations (i.e., A1-B1, A2-B2, B1-C1, B2-C2, C1-D1, and C2-D2) in simultaneity. The trial configurations were the same as those in the sequential phase in terms of sample and comparison presentation.

In the sequential phase, each trial type was presented eight times in total per training block, leading to a total of 16 trials per block of training. Subjects were required to reach a criterion minimum of fifteen out of sixteen correct responses in a block in order to progress to the next relevant block. If subjects failed to reach criterion on eight consecutive training blocks, then the subject was simply moved on to the next block of training. Once subjects completed each of the three sequential blocks of training, they moved on to the simultaneous training phase. In this phase, subjects were tested on each of the six stimulus relations presented in the previous blocks. That is, six trial types in this phase were present: A1-B1 (B2), A2-B2 (B1), B1-C1 (C2), B2-C2 (C1), C1-D1 (D2), and C2-D2 (D1). Stimuli after the dash represent the two comparison stimuli, with the erroneous response option parenthesised. Each of the six trial types was presented in a quasi-random order, with each trial presented once in a cycle of six trials; again, no trial type was presented more than twice in succession. Each trial type was presented three times in each iteration of this phase, leading to each simultaneous block consisting of 18 trials. The criterion for passing this phase was seventeen out of eighteen responses correct on a single block. If the criterion was not achieved, the block recycled. Like the sequential phase, the simultaneous block was allowed to cycle eight times. If eight cycles passed and the subject had not passed the block, the procedure terminated. A subject was **not** considered to have failed the training if they failed to reach criterion in the sequential stage but achieved criterion in the simultaneous stage. However, failure to achieve criterion in the simultaneous stage was considered to be a failure of training. If this occurred, subjects would be thanked for their time, debriefed, and their data omitted from the study. However, all subjects passed the simultaneous phase on training within the eight cycles.

3.2.4.3 Function Acquisition Speed Test

Following the MTS procedure, subjects were required to complete three Function Acquisition Speed Tests (FASTs). The format of the current FAST was identical to that of Experiment 1 (i.e., see 2.2.4.3). The experimental stimuli used in each of the three FASTs varied in terms of the nodal distance between the relevant procedural stimuli. In the 0-node FAST, the stimulus pairs A1-B1 and A2-B2 consisted of the consistent block, while the alternative configuration (i.e., A1-B2, A2-B1) consisted of the inconsistent block. In the 1-node FAST, the relevant consistent stimulus pairs were A1-C1 and A2-C2, while in the 2-node FAST, the consistent stimulus pairs were A1-D1 and A2-D2. The FASTs were presented in a counter-balanced order across subjects in terms of the six possible presentation orders.

3.2.4.4. Stimulus Equivalence Testing

Upon completion of the three FAST procedures, subjects were required to complete a stimulus equivalence testing phase. This testing phase consisted of a match-to-sample procedure format. At the commencement of the procedure, subjects were presented with the following text:

> ***PLEASE PAY ATTENTION WHEN READING THE FOLLOWING INFORMATION***
>
> In a moment some words will appear on this screen. Your task is to look at the word at the top of the screen and choose one of the two words at the bottom of the screen by clicking on it using the computer mouse and cursor.
>
> During these stages the computer will NOT provide you with feedback on your performance. You should use what you learned at the very beginning of this experiment in order to inform how you respond.
> You should try to get as many answers correct as quickly as possible.
> If you have any questions please ask them now.
> When you are ready please click the mouse button to begin the Experiment.

On all trials sample stimuli appeared at the centre-top of the screen in emboldened size 48 Times font, with two comparison stimuli appearing in the lower left and right corners of the screen one second following the presentation of the sample. There was no time limit on responding. Once subjects selected a comparison stimulus, the screen clearer, an ITI of 1s occurred, and the next trial began. That is, no feedback was presented to subjects at any stage of this procedure. The procedure consisted of sixty trials in total, with the relevant stimulus relations from each of the FASTs being tested simultaneously. The trial configurations consisted of A1-B1 (B2), A2-B2 (B1), B1-A1 (A2), B2-A2 (A1), A1-C1 (C2), A2-C2 (C1), C1-A1 (A2), A2-C2 (C1), A1-D1 (D2), A2-D2 (D1), D1-A1 (A2), and D2-A2 (A1), with comparison stimuli to the right of the dash, and the erroneous response option parenthesized. The twelve trial types were presented five times each in quasi-random order, with each trial presented once per twelve trials, and no trial presented twice in a row at any stage. Upon completion of the sixty trials, the procedure terminated, the subject was informed that the experiment was complete, and was invited to ask the experimenter any queries they may have had about the experiment.

3.3 Results

3.3.1 Measurement

Data for the MTS training procedure were recorded for the number of trials of training required to reach criterion on each of the sequential blocks, as well as on the simultaneous block. The mean response times of each block were also recorded. Fluency difference scores on the FAST were calculated using the two metrics outlined in 2.3.1 (i.e., the FAST score, and the rate-fluency differential; RFD). Values were also taken for the accuracy of the subjects on each block, as well as the mean response time of subjects on each block. Data for the MTS testing procedure were recorded in terms of the number of correct responses to each probed stimulus relation group (that is, A1-B1 and A2-B2 were pooled as A-B, etc.) out of 10, as well as the mean time taken to respond on each of these relations.

3.3.2 Descriptive Statistics

3.3.2.1 Stimulus Relations Training

Initial descriptive analyses discerned no major trend in differences in the number of cycles required, nor the mean response time required, to pass sequential MTS training across each of the three sets of stimulus relations. All subjects passed the simultaneous phase of training within the eight-cycle limit imposed. One subject, subject 10, failed to achieve criterion in the A-B sequential training block, but achieved criterion in the B-C and C-D sequential training, as well as the pooled simultaneous training (see Table 5).

Table 5. Table showing the number of block cycles needed by each subject in the sequential and simultaneous MTS training stages (‘Cycles’), as well as the mean response times (‘RT’) for each subject in each stage. Mean response times are expressed in milliseconds, and are rounded to two decimal places.

	MTS Training Phase							
	A-B		B-C		C-D		Simultaneous	
Subject	Cycles	RT	Cycles	RT	Cycles	RT	Cycles	RT
1	1	1167.25	1	1087.94	1	1266.38	3	1628.78
2	2	2337.22	2	1655.34	1	1623.75	7	1935.97
3	1	1286.38	2	1024.00	1	1161.31	1	1358.22
4	1	1529.50	1	1608.81	1	1554.75	1	2078.72
5	1	1493.31	1	1856.94	1	2027.69	1	2905.28
6	1	1564.44	1	1526.25	1	1400.56	2	1672.69
7	1	1976.75	1	2669.00	1	2205.81	1	2049.22
8	2	2985.88	1	1316.44	2	1479.81	1	1510.06
9	2	2451.63	2	1895.09	1	1678.56	3	3360.11
10	8	774.90	2	1227.97	2	1387.33	2	1335.11
11	3	1632.26	1	2727.19	2	3465.38	2	3678.56
12	1	1683.38	1	1310.19	1	1834.56	2	2629.36
13	1	2347.56	1	1786.13	2	1741.84	1	2652.83
14	2	1610.84	1	1473.75	1	1552.63	4	2115.24
15	4*	1228.79	4*	1337.41	4*	1257.44	5	1570.08
16	1	2276.81	3	1622.38	1	1191.94	4	2470.79

*Due to a technical fault, subject 15 continued sequential training even after criterion was met. The subject had achieved criterion after 2 cycles of A-B and C-D, and after one cycle of B-C.

3.3.2.2 Function Acquisition Speed Tests

Descriptive analyses showed that, as expected, FAST scores, RFD scores, accuracy difference scores, and response time differences scores generally decreased in subjects’ scores as the nodal distance of the FAST stimuli increased. Figure 4 shows the changes in RFD scores on each of the FASTs that subjects recorded. Table 6 displays the means for FAST scores and RFD scores for each subject’s three FASTs, while table 7 displays the means for accuracy inter-block differences and response time inter-block differences for each subject’s 3 FASTs.

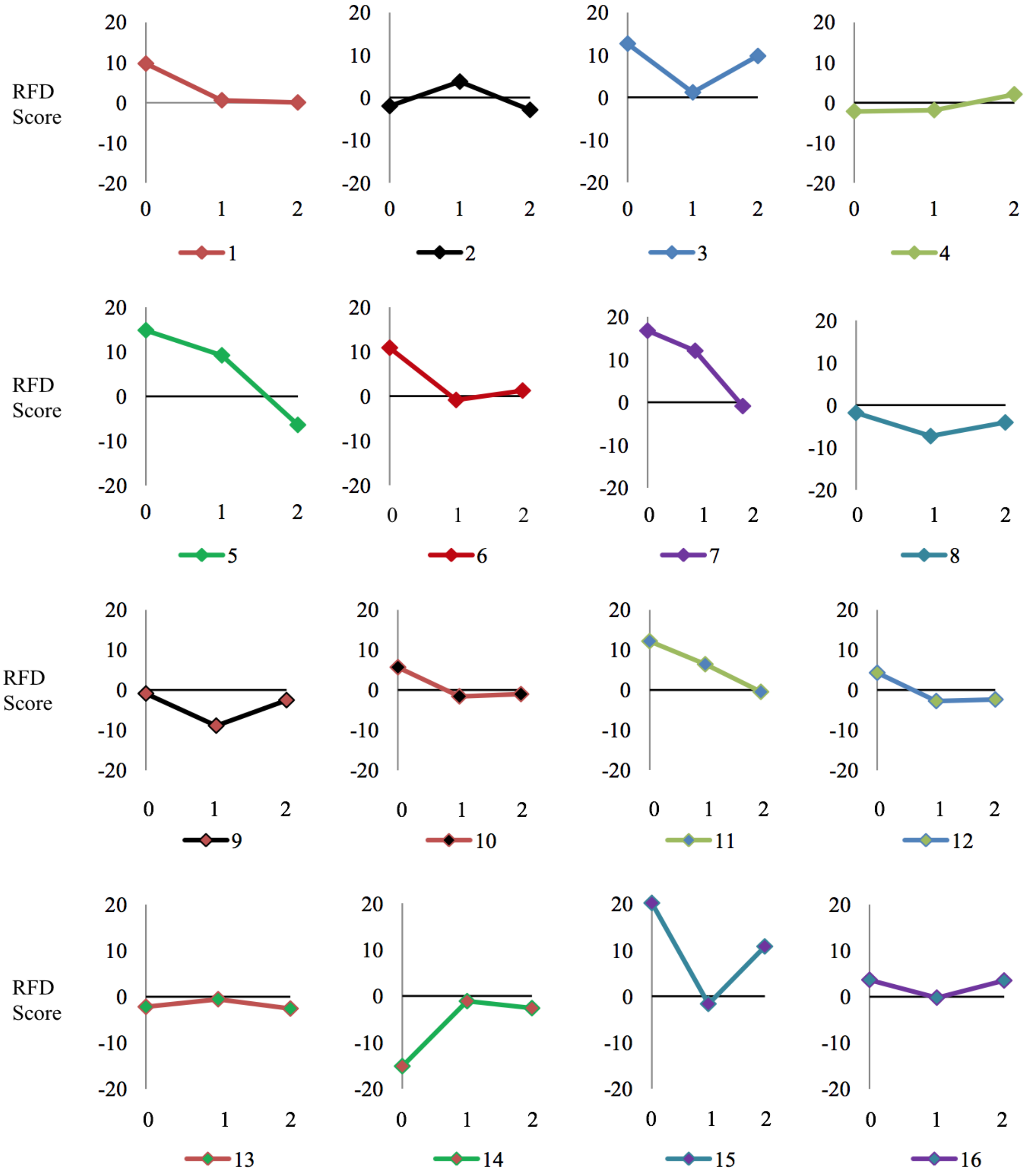


Figure 4 Rate-fluency differential scores recorded for each subject on each FAST. The X-axis values refer to the number of nodes between the stimuli tested in the FAST, while Y-axis values refer to the RFD score recorded in the corresponding FAST. Positive RFD values indicate a more fluent consistent block, while negative RFD values indicate the opposite. The subject number is indicated by legends below X-axis data points (i.e., ‘1’ indicates Subject 1, etc.).

Table 6. Table showing the FAST scores and RFD scores calculated for subjects in each of the three completed FASTs.

	FAST Score			RFD Score		
Subject	0-Node FAST	1-Node FAST	2-Node FAST	0-Node FAST	1-Node FAST	2-Node FAST
1	.14	.04	-.02	9.69	.56	.01
2	-.01	.01	-.01	-1.97	3.77	-2.86
3	.16	0	.08	12.67	1.2	9.77
4	0	-.02	.02	-2.21	-1.89	2.05
5	.17	.11	-.07	14.81	9.19	-6.42
6	.16	-.01	0	10.87	-.86	1.26
7	.18	.17	.01	16.72	12.03	-.91
8	-.05	-.09	-.06	-1.82	-7.31	-4.08
9	.01	-.11	-.04	-.88	-8.89	-2.53
10	.07	-.01	-.01	5.67	-1.59	-1.03
11	.08	.05	-.02	12.15	6.4	-.49
12	.04	-.03	-.01	4.29	-2.76	-2.37
13	-.07	-.03	-.02	-2.2	-.56	-2.58
14	-.15	-.01	-.01	-15.14	-1.11	-2.62
15	.16	.01	.08	20.20	-1.63	10.80
16	.01	0	0	3.62	-.24	3.48

Table 7. Table showing the differences in the number of accurate responses and mean response times between consistent and inconsistent blocks of the FAST for each subject in each FAST. Response time scores are presented in milliseconds.

	Total accuracy difference			Mean response time difference		
Subject	0-Node FAST	1-Node FAST	2-Node FAST	0-Node FAST	1-Node FAST	2-Node FAST
1	6	-2	1	-222.59	275.48	-100.40
2	-1	4	-2	-88.14	-14.66	-91.30
3	9	0	8	440.78	114.50	134.94
4	-2	-1	2	14.96	-66.54	-26.64
5	11	6	-6	396.76	310.16	17.00
6	8	0	1	279.94	-79.30	22.68
7	11	9	-2	562.28	356.58	143.76
8	-1	-6	-3	-115.98	-262.50	-283.06
9	-1	-8	-2	16.70	-386.16	-35.12
10	4	0	0	170.98	-150.28	-85.06
11	10	5	2	211.58	74.80	-275.18
12	3	-1	-1	90.18	-157.58	-116.72
13	-2	1	-1	-2.92	-203.60	-108.48
14	-13	-2	-2	-224.82	102.92	-22.2
15	17	-2	9	132.18	108.20	18.84
16	3	-1	4	39.88	135.54	-73.68

3.3.2.3 Stimulus Equivalence Testing

Data from the MTS testing phase was taken in terms of the number of correct responses out of ten that subjects achieved for each of the six stimulus trial configurations (A-B, B-A, A-C, C-A, A-D, D-A), and the mean response time taken on each of these six configurations. A-B and B-A trial data for subjects 1, 3, 4, and 5 was not differentiated by trial type (i.e., all trials were recorded as ‘A-B’) due to a technical error, and so was not recorded. Per typical criteria from stimulus equivalence studies (Fields & Moss, 2007), it was agreed that subjects who achieved a minimum of nine out of ten responses correct for A-C, C-A, A-D, or D-A had formed stimulus equivalence relations for the relevant trial type (bearing in mind that A-C stimulus equivalence does not necessarily entail C-A responding). According to the equivalence testing criteria, nine out of 16 subjects formed A-C equivalence relations, eight out of 16 subjects formed C-A equivalence relations, eight out of 16 subjects formed A-D equivalence relations, and eight out of 16 subjects formed D-A equivalence relations (Table 8).

Finally, there appeared to be a correspondence between scoring more highly on the consistent block compared to the inconsistent block of the FAST and whether relevant equivalence/discriminative relations were detected in the MTS testing. For this analysis, the stimulus relations were pooled independent of their directionality (i.e., A-B and B-A trials are taken as ‘A-B’, A-C and C-A as ‘A-C’, etc.). Equivalence responding was taken as present when subjects achieved a minimum of 18 out of 20 correct responses for the pooled trial type. Table 9 illustrates this data.

Table 8. Table displaying the total number of correct trials out of ten (‘Cor.’) for each of the six trial types in the MTS testing procedure for each subject, as well as the mean response

time (in milliseconds; RT) of each subject on each trial type. N/A refers to technical errors in the recording of responses.

	Trial Type											
	A-B		B-A		A-C		C-A		A-D		D-A	
Subject	Cor.	RT	Cor.	RT	Cor.	RT	Cor.	RT	Cor.	RT	Cor.	RT
1	NA	NA	NA	NA	3	2819.4	2	2733.2	9	2945.2	9	1838.6
2	10	1680.5	10	1982.4	7	4042.7	9	3432.3	10	2309.3	8	2275.6
3	NA	NA	NA	NA	10	797.7	10	810	10	1064	9	808
4	NA	NA	NA	NA	10	1872.1	10	2208.5	10	2012.4	10	1907.3
5	NA	NA	NA	NA	10	1571.1	10	1549.8	8	2962.6	10	3513.8
6	10	1428.7	10	1815.5	9	2639.7	7	3011.8	5	3510.4	5	3105.4
7	10	1103.4	10	2076.0	10	1195.3	10	1296.1	9	3039.1	10	2275.3
8	6	4606.4	5	5634.3	3	3333.5	2	4388.5	2	3534.3	1	6187.6
9	10	2297.5	9	3151.9	10	7377.6	6	4684.6	4	8540.6	4	5348.6
10	6	726.8	7	2101.2	4	1269.5	4	1256.2	6	1114.3	5	1005.6
11	10	1207.3	10	1927.2	10	5138.1	10	1847.1	10	1980.9	10	2009.2
12	9	4625.8	8	4776.7	5	3014.2	5	3864.4	5	3470.3	0	5174
13	10	1535.8	10	1496.9	10	2040.4	10	1444.6	10	2045.4	9	2352.4
14	6	2130.5	5	4144.4	5	2418.4	5	2040	3	1796.8	7	2693.9
15	9	1693.7	5	1747.3	6	1891.1	2	2368.9	2	2810.8	5	1935.4
16	9	2308.1	10	2138.5	0	2923.1	4	3129.6	2	3656.5	5	2906.9

Table 9. Table comparing whether a positive ('POS'; i.e., consistent block performance superior to that of the inconsistent block) or negative ('NEG'; i.e., consistent block performance inferior to that of the inconsistent block) FAST score/RFD score was seen on each FAST, and whether subjects had formed equivalence relations for the corresponding stimuli in each case in the MTS testing phase ('PASS' refers to the presence of equivalence responding, while 'FAIL' refers to the absence of equivalence responding). Green colouring indicates that FAST Scores, RFD Scores, and MTS testing showed the same score; blue colouring indicates that the MTS testing indicated equivalence where FAST Score/RFD Score showed no effect; red shaping indicates where RFD Scores and FAST Scores showed an effect, but MTS testing did not; orange colouring indicates that the one of the FAST Score/RFD Score was negative, while the other was positive.

	Stimulus Relation								
	A-B			A-C			A-D		
Subject	FAST	RFD	MTS	FAST	RFD	MTS	FAST	RFD	MTS
1	POS	POS	FAIL	POS	POS	FAIL	NEG	POS	PASS
2	NEG	NEG	PASS	POS	POS	FAIL	NEG	NEG	PASS
3	POS	POS	PASS	POS	POS	PASS	POS	POS	PASS
4	POS	NEG	PASS	NEG	NEG	PASS	POS	POS	PASS
5	POS	POS	PASS	POS	POS	PASS	NEG	NEG	PASS
6	POS	POS	PASS	NEG	NEG	FAIL	POS	POS	FAIL
7	POS	POS	PASS	POS	POS	PASS	POS	NEG	PASS
8	NEG	NEG	FAIL	NEG	NEG	FAIL	NEG	NEG	FAIL
9	POS	NEG	PASS	NEG	NEG	FAIL	NEG	NEG	FAIL
10	POS	POS	FAIL	NEG	NEG	FAIL	NEG	NEG	FAIL
11	POS	POS	PASS	POS	POS	PASS	NEG	NEG	PASS
12	POS	POS	FAIL	NEG	NEG	FAIL	NEG	NEG	FAIL
13	NEG	NEG	PASS	NEG	NEG	PASS	NEG	NEG	PASS
14	NEG	NEG	FAIL	NEG	NEG	FAIL	NEG	NEG	FAIL
15	POS	POS	FAIL	POS	NEG	FAIL	POS	POS	FAIL
16	POS	POS	PASS	NEG	NEG	FAIL	NEG	POS	FAIL

3.3.3 Inferential Statistics

3.3.3.1 Frequentist Analyses

Inferential analyses were run in order to determine whether there was any significant effect of nodal distance on the scores (FAST score, RFD score, accuracy difference, and mean response time difference) participants recorded on each of the three FASTs. Scores for all four measures on each of the three FASTs were normally distributed (.2 > *p*s > .055). As such, four one-way repeated-measures ANOVA was run to determine whether FAST scores, RFD scores, accuracy differences, and mean response time differences changed as a consequence of the differential nodal distance in each FAST completed by subjects. For FAST score, the assumption of sphericity was upheld (*p* =.571). As expected, a significant change was noted for FAST scores across the 0-node (*M* = .058, SD = .1), 1-node, (*M* =.005, SD = .067), and 2-node (*M* = -.0047, SD = .041) FASTs, (*F* (2, 14) = 5.207, *p* = .011; η_p^2 = .258, observed power = .79). The assumption of sphericity was also maintained for RFD score (*p* = .89). A significant change was seen for RFD score across the 0-node (*M* = 5.405, *SD* = 9.203), 1-node (*M* = .395, *SD* = 5.385), and 2-node (*M* = .095, *SD* = 4.645) FASTs, (*F* (2, 14) = 4.57, *p* = .019; η_p^2 = .234, observed power = .732). For accuracy differences, the assumption of sphericity was also upheld (*p* = .949). Contrary to what had been conjectured, no significant change was noted for accuracy scores across the 0-node (*M* = 3.87, SD = 7.238), 1-node, (*M* =.13, SD = 4.272), and 2-node (*M* = .5, SD = 3.916) FASTs, (*F* (2, 14) = 3.202, *p* = .055; η_p^2 = .176, observed power = .568). Finally, for response time differences, the assumption of sphericity was maintained (*p* = .34). A significant change was noted for response times across the 0-node (*M* = 106.361, SD = 229.237), 1-node, (*M* =9.848, SD = 212.114), and 2-node (*M* = -55.039, SD = 117.28) FASTs, (*F* (2, 14) = 4.736, *p* = .016; η_p^2 = .24, observed power = .748).

An inferential analysis was run in order to determine whether the sequence in which the FASTs were presented affected either FAST scores or RFD scores on any of the three FASTs. To investigate this, six one-way ANOVAs were run. Each ANOVA had presentation sequence as IV, with either 0-node FAST score, 0-node RFD score, 1-node FAST score, 1-node RFD score, 2-node FAST-score, and 2-node RFD score as DV. The analyses revealed no significant effect of presentation sequence on any of the six DVs (.553 > *p*s > .184). A further analysis was run in order to determine whether any order effects were present for block presentation in each of the three FASTs for either FAST score or RFD score. As such, six independent sample t-tests were run (i.e., two tests for each FAST, one with FAST score as DV, and one with RFD score as DV). No significant effect of FAST block order was found on the 0-node FAST for either RFD score (*p* = .91) or FAST score (*p* = .337). For the 1-node FAST, there was also no significant order effect for RFD score (*p* = .978) or FAST score (*p* = .992). This was also the case for the 2-node FAST for both RFD score (*p* = .905) and FAST score (*p* = .96).

3.3.3.2 Bayesian Analyses

Four Bayesian repeated-measures ANOVAs were run in order to assess the likelihood of an effect present for RFD score, FAST score, accuracy differences, and response time differences, given the data. For FAST score, $BF_{10} = 5.64$, indicating that the data were more indicative of an effect being present than not. For RFD score, $BF_{10} = 3.93$, indicating a similar trend as FAST score. This was also the case for mean response time differences, with $BF_{10} = 3.88$, indicating a greater likelihood of an effect than no effect. No such effect was seen for accuracy differences, $BF_{10} = 1.61$.

3.3.4 Summary

In the Frequentist analyses, a decrease in FAST scores, RFD scores, and mean response times was seen as the nodal distance between FAST stimuli increased. No such effect was no noted for accuracy differences, however. As well as this, there was no confounding effect of FAST sequence presentation, or FAST block order, on these effects. The Bayesian analyses showed similar trend to those of the Frequentist analyses, again with the exception of accuracy differences, which showed no effect as a consequence of the nodal distance in the FASTs.

3.4 Discussion

The current experiment intended to investigate the effect of differential intra-class stimulus relatedness on subsequent scores in the Function Acquisition Speed Test. Each subject was trained to establish two 4-member stimulus equivalence classes. The subjects then completed three FASTs in a randomized order. Each FAST consisted of the formerly-trained stimuli, but differed in terms of their nodal distance (i.e., A-B, A-C, or A-D). After the subject completed the three FASTs, they were exposed to a MTS testing phase, which is a standard means of assessing whether stimulus equivalence relations have been formed by subjects. It was expected that RFD scores, FAST scores, accuracy differences, and response time differences in each FAST would decrease a function of increasing nodal distance between the experimental stimuli. Further, in accordance with the conjectured relationship between FAST Score/RFD Score and accuracy/response time discussed in 2.4, it was expected that if response times showed greater effects than accuracy, then RFD Score would show a greater effect than FAST Score (and vice versa). Finally, it was expected that there would be some degree of correspondence between whether subjects passed or failed equivalence testing of each stimulus relation, and whether the subject showed effects on the FAST which corresponded to that stimulus relation.

The finding of significant changes on RFD Score, FAST Score, and response time differences in accordance with the nodal distance of stimuli in the FASTs was in line with expectations; however, the lack of significant changes in accuracy differences according to this factor was not in accordance with what was expected. This may be the consequence of the relationship between the response window instantiated in the FAST procedure and the nature of responses registered. That is, in general, the presence of a narrow response window punishes slow responding, and therefore results in the narrowing of the variance in response times on both task blocks. However, the consequence of this is the exacerbation of accuracy differences between blocks, given that more difficult trials tend to yield more erroneous responses when response times are constrained (Bolsinova, de Boeck, & Tijmstra, 2016). In a similar fashion, the lengthening of response windows ought to therefore reduce variance in accuracy of scores, while simultaneously exacerbating the effect of response time. It is arguable, therefore, that in the current experiment, the response window was lengthy enough so as to reduce the variance in accuracy of responding, while therefore increasing the variance in response times. Notably, while this accuracy-response time conditional dependency is known, the precise ‘weighting’ that a given response window places on response time or accuracy in the FAST is currently unknown. While this point will be explored in greater depth in Chapter 4, it is notable that the current experiment suggests that a 3 second response window may render pure accuracy differences relatively impotent in detecting differences in the relatedness of experimental stimuli.

Further to the point on the relative impotence of accuracy to response time in the current experiment, it is notable that FAST Score exhibited greater effects sizes in both the Frequentist and Bayesian analyses. This finding was not in line with what had been conjectured following the results of Experiment 1. That is, it had been conjectured that a larger effect size of response time relative to accuracy would therefore lead to a larger effect size of RFD Score relative to FAST Score, given the respective mathematical configurations of the two metrics. Interestingly, the disparity between accuracy and response time effect sizes was even greater in the current experiment than Experiment 1, suggesting that the relationship between FAST/RFD Score and accuracy/response time is not as simplistic as had initially seemed the case. However, the shift from Experiment 1 to Experiment 2 in the relative potency of RFD Score versus FAST Score in detecting differences in relatedness highlights the need for a comprehensive understanding of the psychometric nuances of these measures, a point which will be explicated on in greater detail in Chapter 4.

Notably, the majority of subjects who passed the stimulus equivalence testing phase for a given trial type also showed positive scores on the FAST which used the stimuli of the relevant trial type. There were eight instances where subjects passed MTS testing but showed negative scores on the FAST, which could be used as evidence to suggest that the MTS is in fact more effective in detecting formed stimulus relations than the FAST. However, there were also eight instances where subjects’ scores on the FAST showed positive effects, and the same subjects did not show equivalence on the relevant MTS testing trial. This is of interest to understanding current conceptualizations of stimulus equivalence. Given that former studies (e.g., O'Reilly et al., 2012; 2013) have demonstrated the sensitivity of the FAST to formerly-trained stimulus relations, it is interesting that subjects had learned stimulus equivalence relations to a sufficient degree that transfer of function was present (i.e., in the FAST), but not to a sufficient degree where subjects would reach criterion in the MTS testing. As such, MTS testing trials may not be a sensitive enough metric to detect stimulus equivalence learned in all instances.

Importantly, this point has already been raised within the literature on stimulus equivalence quantification. Fields, Arntzen, and Moksness (2014) demonstrated that the use of a simple card-sorting task was a more sensitive measure of stimulus equivalence formation than traditional MTS testing trials. Specifically, Fields et al. trained subjects in three 5-member stimulus equivalence classes. The 15 stimuli in these equivalence classes were then presented to subjects as individual cards, and subjects were required to sort the fifteen cards into three distinct groups. It was found that in some instances, this sorting task was more sensitive to identifying class formation than standard MTS testing trials. As such, this experiment corroborates the finding that MTS testing trials are not comprehensively sensitive to histories of learning of all subjects. However, in a similar sense, the FAST was also not entirely sensitive to such histories of learning, given the eight subjects who showed equivalence on MTS testing but showed negative scores on the FAST. Broadly speaking, the implication of this is that aspects of stimulus equivalence emerge differentially for individual subjects.

One potential explanation for this may lie in that the procedures are quantifying differing forms of stimulus control. Specifically, from the findings of Experiment 1, it may be suggested that increasing relatedness between stimuli the FAST primarily leads inhibiting effects on the inconsistent block, rather than facilitative effects in the consistent block (given the apparent ceiling effect which is present). It may be argued from this that the FAST effect derives primarily from decreasing scores on the inconsistent block, which therefore be the consequence of increasing S- control (see 2.4). Conversely, it may be argued that behaviour in the MTS testing phase is more sensitive to S+ control, in the sense that the FAST's consistent block may not be sufficiently difficult enough to detect differences in this aspect of stimulus control topography (as formerly discussed in Chapter 2). As such, it may be conjectured that the presence of an effect in the MTS but not in the FAST would indicate the presence of strong S+ control but not S- control. Conversely, the presence of an effect in the FAST but not in the MTS would indicate strong S- control, but not strong S+ control. The confluence of effects in both MTS and the FAST would, by extension, indicate both strong positive and negative stimulus control. In support of this notion are the results of some exploratory post-hoc analyses run in light of this conjecture which were unreported in the thesis thus far (see Appendix E). These revealed that varying nodal distances do indeed lead to significant changes in FAST scores and RFD scores in the inconsistent block, but not in the consistent block. As such, this supports the argument that the effects noted across conditions on the FAST may have been the consequence of changing performances on the inconsistent block, which is arguably a measure of S- control. Therefore it may be that increasing the difficulty of FAST blocks (i.e., presenting tasks in such a way as to globally reduce correct response rates and increase incorrect response rates) would likely aid in the quantification of both S+ and S- control.

Importantly, a further way of experimentally-verifying the aforementioned conjecture would involve the use of a stimulus relations training procedure which can actively control the degree of S+ and S- control which is exerted over S-S relations. While such a procedure is not currently commonly-used in modern stimulus equivalence literature (though see the training procedure used by Johnson and Sidman, 1993), a possible means of achieving this degree of control may be through the use of an adapted Relational Evaluation Procedure (REP; Stewart, Barnes-Holmes, & Roche, 2004). While this will be discussed in greater depth in Chapter 4, it is important to note that exertion of control over the stimulus control topographies of subjects would be of great utility in elucidating the effects of specific degrees and topographies of control in the quantification of stimulus relatedness in the FAST.

The current experiment demonstrated that scores in the FAST changed commensurate to the degree of intra-class relatedness between experimental stimuli. As such, the FAST has demonstrated itself as an effective means of quantifying intra-class stimulus relatedness. Notably, accuracy in the FAST did not differ in accordance with nodal distance, unlike the other three metrics (RFD score, FAST score, and response time). The relatively greater effect size of FAST score to RFD score in spite of the relative potency of response time over accuracy was not in line with expected outcomes, and as such suggests that the relationship between response time and accuracy in these metrics is not as simple as conjectured following Experiment 1. Finally, MTS testing trials and scores in the FAST shared a good deal of concordance, with some degree of divergence in both directions (i.e., the MTS testing trials were sensitive to some stimulus relations that the FAST was not, and vice versa). The remainder of the current these will focus on a greater discussion of the results of both Experiment 1 and Experiment 2.

Chapter 4

General Discussion

4 Discussion

4.1 Research Summary and Main Findings

The current research sought to investigate a fundamental premise of a behaviour-analytic interpretation of implicit psychological measures: that these measures quantify the relatedness between the stimuli used in their procedure. Specifically, it was investigated whether the Function Acquisition Speed Test (FAST) could effectively quantify the differential relatedness of experimentally contrived stimulus classes. Stimulus relatedness was varied in the current studies based on two aspects: via inter-class parameters (i.e., different degrees of training of stimulus relations), or intra-class parameters (i.e., based on the nodal distance between experimental stimuli; Fields et al., 2008). Experiment 1 involved manipulating relatedness along inter-class parameters through the use of iterative differences in match-to-sample training of equivalence classes. Scores in the FAST changed commensurate to the degree of inter-class stimulus relatedness. Experiment 2 manipulated relatedness along intra-class parameters through variations in the nodal distance of stimulus relations used in the FAST procedure. Scores in the FAST generally changed in accordance with the degree of intra-class variance (i.e., greater nodal distance between stimulus relations led to reduced effects in the FAST). As such, the current research has shown the FAST to be a valid measure of stimulus relatedness along both inter-class and intra-class parameters. These findings will now be discussed within the broader research contexts of implicit measures research and stimulus equivalence research.

4.1.1 Quantifying Stimulus Relatedness

As mentioned, the primary goal of the current thesis was to establish the veracity of the Function Acquisition Speed Test as a measure of inter- and intra-class stimulus relatedness. Experiment 1 found that all 4 FAST metrics (FAST score, RFD score, accuracy, and response time) varied in accordance with the degree of inter-class stimulus relatedness of stimuli. Experiment 2, similarly, found that 3 of the 4 FAST metrics (FAST score, RFD score, and response time) varied in accordance with the degree of intra-class stimulus relatedness. Comparing data across the experiments, it is interesting to note that differences between block scores in the FAST were highest when the stimuli were composed of a Real Word set (RFD *M* = 6.37), followed by stimuli which were trained as conditional discriminations (i.e., A-B stimulus relations in Experiment 2; *M* = 5.405), with the next-nearest score being observed for the 3-Iteration condition in Experiment 1 (*M* = 2.3). As such, a single iteration of training for conditionally-discriminative stimulus relations showed a far greater effect in the FAST than 3-iterations of training for 1-node equivalence relations.

This finding has implications for a particular conceptual issue surrounding the implicit measures literature. Specifically, an assumption implicit in the IAT literature is that stimulus category members in general are relatively interchangeable. While it has also been suggested that irrelevant stimulus features may also play a role in informing responses (see De Houwer, 2008), in general, it is assumed that a stimulus' membership within a given stimulus category is sufficient to generate an IAT effect, even if the specific stimuli have not been directly-paired previously in the history of the subject. The current findings suggest that this assumption may be false. That is, the disparity in scores observed between the 3-Iteration A-C stimulus relations FAST in Experiment 1 and the 1-Iteration A-B stimulus relations FAST of Experiment 2 suggest that directly-trained stimulus relations with a less extensive learning history are more readily-detectable by implicit measures than stimulus relations with a more extensive history, but which are based upon a stimulus relation characterized by greater nodal distance. As such, the specific behavioural process which the IAT and other implicit measures quantify may differ depending on the specific dynamics of the learning history the subject has with the specific stimuli (i.e., directly-trained versus when stimuli are members of the same category but have previously not been directly-paired). Consequently, knowing the contribution that nodal distance between stimuli makes to overall implicit test scores ought to inform conclusions regarding the magnitude of implicit test effects.

Notably, the IAT typically uses multiple exemplars of each stimulus category. It may be argued from a behaviour-analytic perspective that this ameliorates the aforementioned difficulty relating to direct training vs. derived relation membership, in the sense that the use of multiple exemplars from stimulus categories may render all of the stimuli functionally-equivalent due to their contrasting with other, distinct stimulus classes (i.e., eliciting inter-class stimulus functions; Moss-Lourenco & Fields, 2011). While this explanation is coherent with previous findings, a counter-argument could be made that the specific dynamics of the learning process that led to the formation of to-be-tested classes, may still impact on implicit tests effect sizes. That is, the extent of inter-class stimulus relatedness results in differences in scores in implicit measures (i.e., Experiment 1). Further, varying the specific dynamics through which stimulus classes are learned results in changes in the degree of stimulus relatedness within these classes (i.e., Experiment 2). If intra-class stimulus relatedness is weaker within sets of stimulus classes, then arguably the classes themselves are less-controlled in the repertoire of the subject, and therefore stimulus relatedness between these classes would also be weaker (cf. the relative disparity between effects in the FAST for briefly-trained A-B conditional discriminations in Experiment 2 compared to extensively-trained A-C equivalence relations in Experiment 1). At the very least, there is ambiguity regarding the effect of specific, differing dynamics of learning on these measures.

An ideal approach to investigating this issue would involve training multiple, extensive stimulus equivalence classes to varying degrees of inter-class relatedness, while also varying the strength of association between stimuli within the classes (i.e., intra-class relatedness), and testing relations using the FAST. Intra-class stimulus relatedness could be manipulated via the format of MTS testing used. For example, a stimulus class of a large number of members, all of which are no more than 1-node apart, could be trained using a many-to-one (MTO) or one-to-many (OTM) MTS training procedure (Hove, 2003). Separately, the linear MTS format used in the current experiments could provide a training procedure with more variable nodal distance (and, as such, intra-class relatedness). Based on the above argument, an interaction between these factors would be expected, in that increasing degrees of intra-class stimulus relatedness should facilitate inter-class stimulus relatedness via improving stimulus control over equivalence classes. For example, it could be expected that scores in a FAST with stimuli from classes with high intra-class relatedness and some given amount of inter-class training should be larger than scores in the FAST with stimulus classes of less intra-class relatedness but slightly more inter-class training. Explication on the specific relationship between inter- and intra-class stimulus relatedness would provide insight into the veracity of the assumption in the implicit literature that stimulus class members are relatively interchangeable in implicit measures.

4.1.2 Conceptualising and Controlling Stimulus Relatedness

An extension of the aforementioned point on potential dynamical interaction between inter- and intra-class stimulus relatedness may be made in terms of the conceptualisation of stimulus relatedness. As mentioned in Chapter 1, stimulus relatedness may be conceptualised simply as a descriptive term referring to the likelihood of stimuli being related in some way. As such, ‘relatedness’ as a phenomenon may refer to a variety of behaviours, such as the probability of stimulus selection (Moss-Lourenco & Fields, 2011), or probability of transfer of function (Fields & Moss, 2007). As also mentioned in Chapter 1, stimulus relatedness can be manipulated along both inter-class and intra-class parameters. However, the vast majority of studies manipulate relatedness using nodal distance effects, and so there is comparably little research which has been conducted in assessing effects of differential inter-class relatedness on equivalence responding. To the author's knowledge, only one published study (Bortoloti et al., 2013) has probed effects on equivalence class strength when inter-class relatedness was varied. Even at this, Bortoloti et al. manipulated inter-class relatedness through differences in the absolute number of training trials presented, rather than in terms of the number of training *iterations* presented (as in the case of Experiment 1). As such, there is a relative dearth of research, not only into the systematic manipulation of stimulus relations along inter-class relatedness parameters, but also in terms of the similarities and differences between manipulations of different aspects of inter-class relatedness. While these issues can only be explicated through empirical research, it is important to note here the lacunae within the stimulus relations literature.

The current thesis has highlighted a third issue which is salient with regard to the conceptualisation and control of stimulus relatedness. Specifically, recall that it was argued earlier (section 4.1.1) that there may be a potential an interaction between intra-class and inter-class stimulus relatedness. Investigation of this potential interaction is salient not only in terms of addressing an implicit assumption in the IAT literature, but also in informing our understanding of the concept of stimulus relatedness. If no interaction between inter-class and intra-class varied relatedness on FAST outcomes were to be seen, then this would imply that these parameters do not involve the same behavioural process. Given that relatedness is generally ill-defined in the literature on stimulus relations, an investigation into these phenomena would provide empirical data which could be used to employ a greater specificity to the precise conceptualisation of stimulus relatedness. If an interaction between inter-class and intra-class manipulated parameters were seen, then this would imply that relatedness is a compound phenomenon which is the product of a number of salient parameters which vary in accordance with one another. If no interaction were noted, then this would instead suggest that relatedness consists of distinct, unrelated parameters. In any case, regardless of whether these parameters interact or not, understanding their relationship is salient to informing a thorough conceptualisation of stimulus relatedness. A thorough conceptualisation is also salient to aiding in the extension of current paradigms for measuring relatedness, allowing for more extensive control of stimulus relations in such research.

4.1.3 Implicit Test Effect Sizes and Comparability of Procedures

A further implication of the current research may be seen in terms of the relationship between implicit test effect sizes and the learning histories of subjects. Specifically, Experiment 1 validated a long-held, but formerly never-tested, assumption of the implicit literature: that implicit measure effect sizes are somehow related to the learning history of subjects with the stimuli used in the implicit procedure (e.g., a larger effect reflects a longer or more effective learning history). Further (as discussed in 4.1.1), Experiment 2 demonstrated that different members of a single stimulus category may elicit differential effect strengths when contrasted with other members, contingent on the specific nature of their membership in the stimulus category to the other relevant stimuli (i.e., previous direct pairing, or equivalence relation). These studies together provide some empirical elucidations of the meaning of implicit test measure scores, as well as representing the first investigations into a stimulus relatedness account of such effects.

Notably, there is some difficulty in generalizing the current findings to other implicit measures. In particular, there are a number of key procedural differences between the FAST and other implicit measures (cf. Table 2) which may impede the veracity of inferential generalisibility from the current findings. For example, recall that the FAST measures subjects' acquisition of specific functional responses between pairs of stimuli, the pairings of which are juxtaposed across two blocks. Further, recall that the IAT measures subjects' speed of responding (using a *D* algorithmic transformation) across the fourth and seventh block of the procedure, both of which are preceded by a number of training blocks requiring minimum criteria for speed and accuracy responses. By contrast, the IRAP measures subjects' speed of responding using the same *D* transformation as the IAT, but employs between 1 and 4 practice blocks (until subjects reach speed and accuracy criteria), and then 3 critical test block pairs. Given the employment of practice/preliminary blocks in the IAT and IRAP, it may therefore be suggested that these measures do not measure the acquisition of functional response classes, but rather, measure rates of responding *after* the initial acquisition of the responses has satisfied the arbitrary acquisition criteria of the preliminary blocks. It may be difficult therefore to generalize findings from these FAST studies to the IAT and IRAP, as the FAST may be indexing behavioral dynamics early in the acquisition phases of the IAT and IRAP rather than the dynamics of behavior after it has stabilised.

As well as this, the metricisation of effects differs across these procedures, as formerly-noted. While the FAST employs an amalgam of response time and accuracy, the IAT and IRAP employ a *D* transformation based on response time scores only. Given that these procedures index behavioral dynamics at different stages of a somewhat generic implicit testing process, and metricize these measures differently, our ability to generalise the current findings in any meaningful way across procedures is limited. This broader point ought to be borne in mind by researchers in the field of implicit measures generally, as the relative comparability of procedures is likely variant. As such, this underlines the need for paradigms utilising alternative implicit measures to replicate findings from other measures. Importantly, noting similarities and differences in effects across these procedures would provide salient information regarding the relative comparability of such measures. However, the current findings at the very least suggest that the FAST procedure is an effective measure of differential relatedness.

4.2 The FAST and Single-Subject Measurement

Traditionally within the experimental analysis of behaviour, the veracity of procedures is gauged at the within-subject level. That is, an idealised behaviour-analytic measure of differential stimulus relatedness ought to effectively discriminate differences at a descriptive, individual level, without requiring the use of inferential statistical procedures. For example, if a child is unable to read, and the administration of a behavioural intervention leads to the child being able to read, an inferential statistical procedure is not required to determine whether there is a difference in performance before and after the intervention (cf. Sidman, 1960). In a similar way, the FAST procedure ideally should provide information about an individual subject's learning history with relative precision. Experiment 2 represented a step in the direction of this form of within-subjects measurement. However, the FAST did not show pronounced effects for all subjects in Experiment 2. In particular, Subjects 2, 4, 8, 9, 13, and 14 did not show any effect on the FAST for A-B relations, while Subjects 3, 15, and 16 showed a greater effect on the FAST for A-D relations than A-C relations (Figure X). In spite of this, Subjects 1, 5, 6, 7, 10, 11, and 12 showed the exact trend of effects expected at the commencement of the experiment. As such, the FAST shows utility as a within-subjects measure of relatedness, but in Experiment 2 this utility was only demonstrable in 7 of 16 subjects. This is much greater than chance level (i.e., six permutations of relative effect size were possible; 0-1-2, 0-2-1, 1-2-0, 1-0-2, 2-0-1, 2-1-0, where the sequence of numbers represents decreasing effect sizes), but still not at the level of specificity required by experimental behaviour analysis. Some issues around the development of the FAST as a within-subjects tool, and the ameliorative steps to be taken, will now be discussed.

4.2.1 Effects of Order

In both Experiment 1 and Experiment 2, there were no notable effects of block order in either between- or within-subjects inferential analyses. As well as this, in Experiment 2, there was also no effect of order of presentation of the different FASTs. A cursory viewing of the data would therefore suggest that block order and FAST presentation order are irrelevant features. While this is true at an inferential level, where even within-subjects analyses are conducted in terms of mean scores, it is not applicable at an individual level. Consider Subject 14 in Experiment 2, for example. The subject showed a large, negative RFD score for the A-B FAST, and no notable effect for either the A-C or A-D FAST. This would suggest that the subject found it easier to pair inconsistent stimuli in the A-B FAST than to pair consistent stimuli: an effect for which the conceptual assumptions for this experiment would be unable to account. Notably, the subject completed the A-B FAST first, and completed the consistent block of this FAST first. In looking at the raw data for this subject, the first 10 responses registered consisted of 8 incorrect responses and 2 correct responses (i.e., below even a chance level of responding). This very low level of accuracy would suggest that either the subject genuinely found pairing the stimuli inconsistent to their learning easier for as yet unknown reasons related to unidentified stimulus control features, the subject was responding in opposition to previous training purposefully, or the subject’s difficulty in responding correctly is the consequence of unfamiliarity with the response format of the FAST. We will now consider each of these three possibilities in turn.

Given that there is no known stimulus control reason why the subject should find it easier to learn the formation of functional response class containing relationally inconsistent stimuli, it is at present a rather speculative explanation for the observed effect. If, on the other hand, the subject is able to purposefully respond in opposition to their previous training, either the MTS training is ineffective, or the FAST procedure is unreliable because the acquisition rates measured by the test are not consistently affected by previous learning histories. It is unlikely, however, that the training is ineffective, given that such MTS training is regularly used to train S-S relations, and has been used with great utility in a number of studies in training stimulus relations (e.g., Arntzen & Holth, 1997; Bortoloti et al., 2013; Fields & Moss, 2007; Moss-Lourenco & Fields, 2011). Finally, if the procedure can be readily faked, then the argument can be made that the FAST is not a very good implicit measure (De Houwer et al., 2009). While an empirical analysis into the fakeability of the FAST has not been conducted to date, previous analyses using other implicit procedures have shown that fakeability is relatively difficult to achieve (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007; Steffens, 2004), though possible in some instances when subjects are instructed on how to do so (Fiedler & Bluemke, 2005; Hughes et al., 2016). In particular, faking has been found to be especially difficult for subjects on the first run of an implicit procedure (Fiedler & Bluemke, 2005), and so it is unlikely that faking is the source of the effects seen for Subject 14.

It might be argued that the most likely cause of this large, negative score in Subject 14’s A-B FAST performance is due to unfamiliarity with the response format of the FAST. The difference in acquisition rates recorded for Subject 14 across the consistent and inconsistent blocks of the A-B FAST may therefore not be due to differences in relatedness *per se*. Although subjects are required to learn a new functional response configuration on both blocks of the FAST, the subject is only required on the first of these blocks to also acquire the skills relevant to the completion of the procedure, such as responding within the response window[4], responding using the specific key presses, etc. This may partially explain the use of initial practice blocks in the IRAP and IAT procedures. That is, such practice

---

[4] Note that for many subjects in the FAST (including Subject 14), it is common that the first response on the first block presented is a timeout, due to the fact that subjects are not explicitly instructed of the presence of a response window prior to commencement of the procedure. This timeout can sometimes extend as far as the third response before subjects’ responses are brought under temporal control.

blocks aid in bringing responding under control of the stimulus presentation contingencies prior to testing. While the instantiation of such practice blocks also leads to complex conceptual issues, their use in the FAST might improve procedural familiarity, and hence remove this confound in the effort to develop a single-subject FAST. However, notably, providing practice prior to the critical phase of the procedure may also impede the acquisition differential between blocks. More specifically, if subjects achieve a high level of fluency in the operant established by the procedure, this would make differences in learning across main test blocks harder to discriminate. The amount of initial practice which can provide subjects with procedural familiarity, while also not confounding the effects seen in the procedure, should be determined experimentally, and is a salient issue to address in the development of a single-subject FAST.

4.2.2 Effects of Practice

Practice effects (i.e., improvement in performance on the FAST procedure as a consequence of having formerly completed the procedure) were not discernible at any group level for the inferential analyses conducted in Experiment 2. However, there may be instances in which practice may impede the presence of effects. Consider Subject 2 from Experiment 2, for example. The subject received the A-D FAST first, then the A-C FAST, and finally the A-B FAST. As expected, the scores on the A-D FAST were lower than scores on the A-C FAST. However, scores on the A-B FAST were lower than scores on the A-C FAST, and showed no effect in either direction. It is of interest that the subject showed the initial expected trend from A-D to A-C, but failed to exhibit an effect for A-B. It may have been the case for the A-B FAST that the subject was already sufficiently fluent in the FAST procedure, having already completed two FASTs, and so failed to show any discernible difference in block scores across the A-B FAST, which would otherwise be expected to yield the largest effect. Interestingly, this subject achieved 20 out of 20 correct responses in the

MTS testing phase for A-B relations, which is indicative of equivalence relations having been well-formed. This may indicate that there is an effect of practice in the FAST for some subjects, which is a salient issue to address in the development of a within-subjects FAST procedure.

The issue of practice effects is widely-noted, both within the implicit measures literature (Bluemke & Fiedler, 2009; Röhner, Schröder-Abé & Schütz, 2011) and in the field of psychological testing generally (Hausknecht, Halpert, Di Paolo & Moriarty Gerrard, 2007). However, investigation into, and amelioration of, this issue is highly relevant to the development of an effective within-subject measure of stimulus relatedness. One experimental design in particular which could assist in understanding the effect of practice would involve the administration of multiple FASTs using either arbitrary stimuli or real word stimuli of known relatedness. These FASTs could contain either the same or different words in each iteration (when arbitrary stimuli are used), or either the same words, synonyms of the same words, or different words (when real words are used). The effects of practice should be seen at some level for each of these conditions. However, such a design would also allow for the examination of the interaction between practice and other test aspects, such as procedural familiarity and stimulus familiarity. Such investigations into the FAST procedure are pertinent in order to understand and assess the suitability of the FAST as a single-subject measure of stimulus relatedness. However, given the greater-than-chance measurement of differential FAST scores commensurate to relatedness noted in Experiment 2, the current work provides evidence for at least the potential suitability of the FAST as a within-subjects measure.

4.3 Quantifying Fluency in the FAST

A secondary aim of the current research was to compare the relative utilities of the typical FAST fluency quantification metric (the FAST score), and a new metric developed by the author based on a Skinnerian notion of fluency; the rate-fluency differential (RFD) score. In Experiment 1, the RFD score was more sensitive to differences in experimentally-controlled inter-class stimulus relatedness than the FAST score, in terms of both Bayesian and Frequentist analyses. It was conjectured that the cause of this was due to the greater effect of response time compared to accuracy noted in Experiment 1, and the purported differences in the weighting of these values in the RFD score FAST scores. However, in Chapter 3, this conjecture was challenged. That is, the FAST score was seen as more sensitive than RFD score in measuring intra-class stimulus relatedness in Experiment 2, in spite of the fact that response time was again more sensitive than accuracy. The simplistic reduction of differentially proportionate weighting of accuracy and response time in these measures is therefore likely not the sole explanation for these differences.

It is beyond the scope of the current thesis to engage in a psychometric evaluation of the two measures. As such, future studies specifically investigating these measures (and potential other alternatives) are required in order to draw a full evaluation of the utility of these approaches to quantification. However, it is notable that the FAST score is somewhat more conceptually-opaque than the RFD score. More specifically, recall that the RFD score is the result of subtracting the response rate differential (RRD) score of the inconsistent block away from that of the consistent block. Also recall that the FAST score is computed through subtracting the block-slope score of the inconsistent block from that of the consistent block. Conceptually, the RRD score for each block is relatively transparent: it simply reflects the difference in the rate of correct responding and the rate of incorrect responding. If random responding constitutes responding correctly and incorrectly at equal rates, then an RRD score for a given block will be zero. This method of scoring thus effectively circumvents the issue of detecting random responding, which has proven difficult to avoid in social psychological research (Osborne & Blanchard, 2011). By contrast, random responding scored through the use of the FAST score can yield block-slope scores which range between .25 and .33 approximately (based on anecdotal testing by the author). Given that block-slope scores typically range between .25 and .46, randomly responding may not be easily discerned from responding under the stimulus control of the procedure. As such, the block-slope approach does not as readily circumvent the issue of random responding as the RRD score. As well as this, the block-slope score is in some ways conceptually-opaque. The use of a cumulative record (which the FAST score employs) is well-established in behaviour analysis as a robust tool for the analysis of rates of responding (Skinner, 1959). However, fitting a regression line to this record (the process by which block-slope scores are calculated) adds a deal of conceptual opacity, in the sense that it is difficult to precisely define what is being measured (given that the score is in some way mathematically-abstract).

A further point of interest regarding comparison of these two metrics is in terms of their ability to facilitate metrical comparisons across different formats of the FAST. Take, for example, an instance in which FAST blocks are lengthened to 100 trials instead of 50. Assume that a researcher wishes to make a comparison between scores on a block from this FAST and a block from a previous FAST which followed the typical 50 trial block length. Using the RRD score for the blocks would facilitate a relatively straightforward comparison, in that differences in *rates* of correct and incorrect responding are being measured. Assume, for example, that a subject takes the same length of time to respond on each trial (e.g., 1000ms per trial). Further, this subject responds incorrectly on two trials in the 100-length block, and incorrectly on one trial in the 50-length block. Given this information, the *rate* of incorrect responding calculated for both blocks would be the same (i.e., .48 incorrect responses per minute including ITI and feedback times in the calculation). Due to this reliance on rates, the RRD score for both blocks would be the same (RRD = 23.04). By contrast, the FAST score varies slightly under the same conditions when the erroneous response(s) are placed at the beginning of each FAST block, with the 100-trial block interpreted as a slope score of .39976, and the 50-trial block interpreted as a slope score of .4. While this slight difference would arguably be essentially non-discernible in any analysis, it is important to note that the impact of individual erroneous responses is affected minimally by the overall length of the procedure, *even if* the rate of erroneous responding is the same on both blocks.

Further, when the erroneous response(s) is/are occur at the end of the blocks (i.e., rather than at the beginning), the slope score of the 50-trial block is .39905, while the slope score of the 100-trial block is .39929. Note here not only the disparity in the two scores in spite of the identical rate of responding, but also the difference in these scores compared to those in the preceding paragraph, where erroneous responses were placed at the beginning, rather than the end, of the block. The FAST score is affected not only by trial length of blocks, but also where precisely erroneous responses are recorded within the block. As such, this variability of the FAST score arguably makes it less suitable than RFD score for comparing scores on variants of the FAST procedure. While the argument may be made that the relative disparities in these different permutations are so slight that outcome comparisons across procedures may still be readily facilitated by FAST score, it is important to note that the convergence of multiple, seemingly-benign factors can compound to complicate the process of comparing outcomes from studies using different procedures (cf. the logic of Meehl, 1967). Rather than deal with this issue by simply formalising the test format (and therefore potentially reifying the constructs the test purportedly measures, as has arguably been done by the IAT and IRAP), the more functional approach to this issue is to understand empirically how each variation in the procedure and effect quantification methods relate precisely to computed test scores. At present, it is arguable that the employment of the RFD score as a conceptually-transparent metric, which is amenable to comparison across procedures, may be more appropriate in some instances than the FAST score.

In spite of the above discussion, it should be noted that both variations (i.e., FAST score and RFD score) in the quantification of implicit measure effects overcome a salient issue which plagues that of the typical response time scoring of other implicit measures. Specifically, the use of means and standard deviations to analyze response times (as employed by the *D* algorithm) assumes that these values in general are normally distributed, in spite of the fact that response time data instead typically fit an ex-Gaussian distribution (Heathcote, Popiel, & Mewhort, 1991). Treating response time data as normally distributed can then affect overall effect sizes, and consequently lead to the obfuscation of salient differences in response times (Whelan, 2008). Given that neither the FAST score nor RFD score parameterize the response time data, these scores avoid this confound. Notably, however, any analyses run in the current thesis using response time parameterized these scores using means and standard deviations. As such, it may be preferable in future studies using the FAST to examine response times through analyses along ex-Gaussian parameters. As well as this, future studies should seek to assess to what degree findings on other implicit measures are affected by this differential parameterization.

4.4 Quantifying Emergent Stimulus Classes

The findings from the two experiments suggest that the FAST may have utility in measuring the variance in class strength that follows in tandem with the differential emergence of varying components of stimulus equivalence. Equivalence classes are not single behavioral units (Pilgrim & Galizio, 2000). For instance, symmetry is demonstrable before transitivity occurs (Dube, Green, & Serna, 1993), and reversing pre-established symmetry relations does not necessarily reverse the accompanying pre-existing transitivity (Pilgrim & Galizio, 1990). Measuring the emergence and strength of equivalence relations (both in terms of class strength and intra-class stimulus relatedness of nodally-different stimulus relations) at different points in training would aid in providing a comprehensive picture of the behavioral process of stimulus equivalence (Doughty, Brierley, Eways, & Kastner, 2014). A typical method for measuring equivalence is through the use of multiple MTS probe trials in a testing phase (as in the case of the MTS in Experiment 1). However, it has been shown that responding in accordance with equivalence in these testing formats typically requires multiple trials or phases (Doughty, Leake, & Stoudemire, 2014). Equivalence responding has also been observed using entirely untrained stimuli (Harrison & Green, 1990), which indicates that the MTS probe trial context and format may confound the phenomenon it intends to measure via directly facilitating the formation of equivalence relations solely in terms of its format.

As mentioned in Chapter 1, card sorting tasks have been offered as a more sensitive alternative to MTS probe trial formats (Fields, et al., 2014; Fields, Arntzen, Nartey, & Eilifsen, 2012). However, such sorting tasks only answer the discrete question of whether or not relations have formed. The FAST, in contrast offers a more continuous metric that indexes the current level of relatedness of stimuli. It avoids the issue of the potential interference in class formation by the probe measure itself, given that stimuli are never matched in the procedure. It also provides much more comprehensive data than card-sorting, and can be administered multiple times intermittently during the equivalence training due to its short duration. Of course, it may be argued that multiple iterations of the FAST may potentially confound the training contingencies of the MTS training procedure. However, it should be borne in mind that the FAST employs both class-consistent and class-inconsistent contingencies in equal measure and so might not be expected to facilitate or militate against class formation during repeated probe phases.

It might be argued in that the FAST does not provide a definitive indicator of whether or not stimulus relations have been formed, given the continuous nature of its score variable. While this criticism is valid, it is somewhat of a truism, in the sense that all test criteria for determining the formation of equivalence relations are ultimately arbitrary (e.g., 90% correct on MTS trials) and such criteria can easily be arrived at for the FAST method through further exploratory research linked to pragmatic criteria associated with training protocols.

4.5 Potential Confounds of the Current Research

4.5.1 Effect of Contrast Stimuli

Experiment 2 manipulated the nodal distance between the relevant experimental stimuli in each FAST, and both stimulus pairs employed in the test were of equal nodal distance. That is, A1 and D1 stimuli were always presented with A2 and D2 stimuli, A1 and C1 stimuli were always presented with A2 and C2 stimuli, and A1 and B1 stimuli were always presented with A2 and B2 stimuli. As such, it is unclear to what degree the presence and relatedness of one pair of stimuli relative to the other set of stimuli (i.e., contrast stimuli) influenced the effects seen in each FAST. For example, it is unclear what effect would be seen if one stimulus pair involved a relation of 0-nodal distance (A1-B1), while the second pair involved a relation of 2-nodal distance (A2-D2). While this is not a confound to Experiment 2 in the sense that it does not impede upon the veracity of the inferences made in the experiment, it would be of interest in future experiments to examine the effect of differences in nodal distances across each pair of stimuli used in the test, as well as the effect of entirely neutral contrast stimuli. For instance, consider two conditions using the FAST. In both conditions, subjects would be trained in the same stimulus relations, and the first pair of stimuli would be the same in both FASTs (e.g., A1-B1). However, in one condition, the contrast pair would be of a different nodal distance to the first pair (e.g., A2-D2), and in the second condition, the second stimulus pair would be of the same nodal distance (e.g., A2-B2). Such an experiment would allow determination of whether the nodal distance of the contrasting stimulus pair relative to the other pair affects scores in the FAST. By the account of Fields and Moss (2007), the stimuli of differing nodal distance should be functionally-equivalent (i.e., nodal distance should be irrelevant) when presented in the context of contrasting with another stimulus class. However, given that Experiment 2 found differences in scores as a consequence of nodal distance in the FAST in spite of this same potential context for functional equivalence within the class, functional equivalence between stimuli is unlikely to be seen.

If differences in outcomes were observed across the above two conditions, it would suggest that intra-class relatedness of one stimulus class can affect the overall scores on the procedure. In particular, it would be of interest to investigate whether, in the above-outlined experiment, fluency scores for A1-B1 pairings on consistent blocks are affected by the presence of differentially-related contrast pairs. This issue has been raised in the context of the IAT literature in terms of the salience asymmetry account of IAT effects (Rothermund & Wentura, 2004). In brief, this account suggests that effects on one block of implicit measures should be enhanced by virtue of the clarity of the contrast category. In behaviour-analytic terms, this may be understood as the contrast stimuli exerting contextual control over the other stimulus pairs. By this reasoning, it should be expected that a more nodally-distal contrast pair ought to reduce scores in the FAST in general, due to the entailed reduction in the salience of the relation, *as well as* specifically reducing the fluency of responding on the consistent block containing the strongly related (i.e., in this example, A1-B1) stimuli.

Given the suspected role of contrasting relations in the formation of stimulus relations and the measurement of relatedness, it may be possible to develop a measure of the absolute relatedness of stimuli, independent of contrast stimuli, through the use of a two-phase FAST (e.g., Cartwright, 2013). To adopt the paradigm of Experiment 2 as an example, this would involve the administration of three two-phase FAST procedures (i.e., six FASTs in total). An A-B two phase FAST would involve a first phase of administration of a FAST using (for example) A1-B1 as a stimulus pair, and two other neutral stimulus categories (e.g., previously unseen nonsense words, or real words of no relation; N1-N2) as the second stimulus pair. Then, subjects would complete the second phase of the A-B FAST, using A2-B2 as a first stimulus pair, and some other neutral stimuli as the second pair. This procedure would also be administered for A-C and A-D stimulus relations. In effect, such a procedure would allow for the measurement of relatedness of the trained stimuli without the issue of potential confounding inter-class relatedness due to the presence of relevant contrast stimuli. As such, it may be argued that this would allow for inquiry into solely S+ control between stimuli, independent of negative stimulus control[5], or vice versa.

The reader might assume that the relatedness between and A1-B1 stimuli should be effectively the same as the relatedness between A2-B2 stimuli, as well as being the case for A1-C1 and A2-C2, and so on. As such, the second phase of each two-phase FAST may be seen as superfluous. However, this second phase could also be adapted to inquire directly into S- control. That is, in this phase, a member of each stimulus class of differing nodal distance could constitute a stimulus pair (e.g., A1-B2 for 0-node, A1-C2 for 1 node, etc.). As in the first phase of the procedure, neutral stimuli would constitute the other stimulus pair. As such,

---

[5] Note that it may be reductive to refer to S+ control and S- control as independent entities (McIlvane & Dube, 2003). Positive and negative stimulus control can interact in exacerbating the frequency/intensity of responses. However, the question of the applicability of this point for FAST research has not been empirically investigated. As such, for the current discussion, relative independence will be assumed (though empirical work in the FAST should be conducted in order to validate/refute this point).

this two-phase FAST procedure would constitute two FASTs in total, each of which would differentially inquire solely into either S+ or S- control. For an A-B FAST, for example, the S+ FAST would consist of an A1-B1 stimulus pair, and an N1-N2 stimulus pair. The consistent block of this FAST would involve pairing A1-B1 and N1-N2, while the inconsistent block would involve pairing (say) A1-N1 and B1-N2. The score in this FAST would provide an absolute measure of S+ control between A1 and B1. The S- FAST would then consist of A1-B2 as a stimulus pair, and N1-N2 as another pair. The ‘consistent’ block would involve pairing A1-N1 and B2-N2 (given that the A1-B2 pair should in fact be more difficult to pair). The ‘inconsistent’ block would involve pairing A1-B2 and N1-N2. As such, this FAST would provide an absolute measure of S- control. The combination of the two phases of this A-B FAST, therefore, would allow for the inquiry into the degree of both S+ and S- control of trained S-S relations, independent of the presence of the other control.

4.5.2 Limits of MTS in Controlling Stimulus Relations

As alluded to in 3.4, the current experiments lacked a degree of experimental control in the sense that the MTS training did not allow for specificity of stimulus control over stimulus relations in terms of S+ and S- control. Further, as alluded to in section 2.4, it may have been the case that in additional iterations of MTS training in Experiment 1, S- control was not as well established as S+ control. Specifically, subjects’ performances improved on the MTS training blocks across successive iterations, which may have resulted in a lack of S-response opportunities in later trials. More specifically, subjects became more fluent in the MTS procedure with more sessions, and consequently were not provided with opportunities to contact punishment for erroneous responses. Hence, subjects’ responses likely increasingly came under S+ control and decreasingly under S- control as training progressed. At the same time, we must be mindful that S- control was demonstrably increasing to at least some extent as training iterations increased because a decrease in fluency on inconsistent FAST blocks was noted across conditions from least to most training.

Even beyond the issues raised in the discussions of the two current experiments, a class training procedure which can independently manipulate S+ and S- control of stimulus relations would be of great utility in research into stimulus relatedness generally, as it would provide a much more comprehensive picture of the stimulus control topographies in operation during training procedures. As briefly mentioned in Chapter 3, an adapted version of the Relational Evaluation Procedure may provide such a tool. The REP involves learning relations between stimuli through responding in accordance with the veracity of propositions about those stimuli (Stewart, Barnes-Holmes, & Roche, 2004). Essentially, this adapted procedure would allow subjects to respond ‘true’ or ‘false’ to specific relations between stimuli. Specifying the relations between stimuli would allow for the explicit establishment of either S- control, S+ control, or both, and would therefore allow for greater control of the stimulus control topographies of subjects. For example, a training format using this procedure could involve the exclusive use of negated statements in order to generate negative-only control over S-S relations (e.g., responding ‘true’ or ‘false’ to “A1 is not the same as B1”).

Importantly, using an adapted version of the REP would require empirical investigation in order to ensure that it was effective in establishing equivalence relations. Little empirical research to date has been conducted in this regard, with one study showing a precursor REP to be slightly less effective than MTS in training stimulus relations (Cullinan, Barnes-Holmes & Smeets, 2000), and another showing it as effective in training analogical derived relations (Stewart et al., 2004). More recently, however, the REP has shown effectiveness as a means of training varying relational configurations to children (Hayes, 2017). An empirical comparison of the MTS with the REP in terms of effectiveness in training stimulus relations would shed light on whether the REP is a viable procedure to exert control over stimulus control topographies in relational learning. However, if not the REP, a procedure capable of exerting such differential S+ and S- control is necessary in order to further advance research into stimulus relatedness and its effects on implicit measures.

4.6 Expanding the Continuum of scores in the FAST

In spite of the purportedly much greater history of learning which subjects had with the real word stimuli used in Experiment 1, it is notable that both the 3-Iteration condition in Experiment 1 and the A-B FAST in Experiment 2 both showed effects in the FAST which were nearly identical to the real word condition. Given the vastly more expansive history of learning that the real word stimuli likely had in comparison to the much shorter history of training provided to subjects in establishing the arbitrary stimulus relations, it is rather curious that these scores do not diverge to a greater degree. It may be conjectured from this that mere cross-sectional measures (such as a single sitting of the FAST) may not have the specificity to measure such differences, as variance in scores is ultimately limited (recall, for example, that scores on the consistent block quickly reach a high level of fluency with comparably little relatedness, and so effects are primarily driven by differing S- control, the range of which is limited). The already-constrained variance in scores is compounded by the addition of intertrial interval and feedback presentation time to cumulative time values during the calculation of the acquisition rate, which together ultimately reduce the relative differences in scores by adding regular time constants to the cumulative record and therefore reducing the differentiating functions of the temporal features of test performance (recall also the discussion regarding scoring in 2.4).

It is apparent that a single iteration of the FAST procedure is not sufficient to capture variance in scores that should be expected at degrees of relatedness beyond cursory iterations of stimulus relations training. As such, the incorporation of a further measure beyond the cross-sectional FAST score/RFD score is required in order to gain greater precision in measuring differential relatedness. One avenue to pursue in the development of such a measure relates to the fact that the resistance of behaviour to change is not measured only in a single iteration of a procedure. Rather, it is also indicated by the persistence (or lack thereof) of the behaviour across iterations. This is also the case for class compatibilities (Nevin & Grace, 2000). It may therefore be argued that a more effective means of scoring these tests would incorporate the temporal persistence of FAST/RFD scores across test iterations, as a further index of relatedness. For instance, multiple sittings of the FAST could be used to measure the rate of extinction of FAST scores as an index of stimulus relatedness. This ‘meta-FAST’ approach would potentially provide a novel and more enriched approach to measuring test effects.

The literature on the IAT arguably supports the notion of the meta-FAST as an index of stimulus relatedness. Specifically, test-retest coefficients for the IAT tend to fall between .3 and .8, with these scores varying in accordance with the conceptual categories used (Nosek, Greenwald, & Banaji, 2007). This suggests that retests on the IAT generally tend to differ relative to original sittings, but the relative amount of difference in scores (i.e., the specific coefficient) is the consequence of the specific stimulus classes used in the measure. This variance due to differing stimulus classes may be indicative of differences in the persistence of specific relations in comparison to others, which (from a stimulus relatedness account) would suggest that different degrees of relatedness are being quantified through the variance in this test-retest statistic. Although empirical research is needed in order to ensure that such changes across sittings are present in the FAST as well as the IAT, the presence of such differences in test-retest coefficients in the IAT literature lends itself to the suggestion that temporal persistence could be a viable extension of the FAST metric (and indeed, other implicit measures generally) in interrogating stimulus relatedness more thoroughly.

A potential barrier to the meta-FAST approach may be that practice effects produced by repeated exposures may result in effects extinguishing due to procedural familiarity, rather than due to extinction of the resistance to behavioural change towards stimulus classes. The earlier-outlined study in 4.2.2 (administering multiple FASTs with differences across conditions in terms of the stimuli used in each FAST, in order to parse procedural familiarity from stimulus familiarity) would provide insight into this potential confound. If practice effects were found to be influential in the extinction of effects, one potential ameliorative tactic would be to simply make the FAST procedure more difficult to begin with. As discussed formerly in both 2.4 and 3.4, the consistent block of the FAST in particular may not be sufficiently difficult enough to discriminate differences in degrees of positive stimulus control beyond very slight degrees at the earliest stages of learning. Increasing the difficulty of the blocks may serve to improve the ability of the consistent block to measure differences in S+ control between stimuli, as well as serving to lengthen the number of runs needed for differences in scores to extinguish (due as to the increased difficulty of the inconsistent block). A number of slight procedural changes could exacerbate the difficulty of FAST blocks. For example, Lakens, Schneider, Jostmann, and Schubert (2011) found that simply having response keys closer together can increase the difficulty of binary categorisation tasks (e.g., the FAST).

A further procedural manipulation which may increase the difficulty of the blocks in the FAST would be to shorten the response window provided to subjects, thus making responding correctly more difficult, and increasing rates of incorrect responding (Bolsinova & Maris, 2015). However, it is important to note that changing the response window implemented within the procedure will also subsequently impact upon the relative differences in the response time/accuracy of subjects. Specifically, the narrowing of the response window would more heavily punish slow responding, and as such, result in generally quicker responding across both blocks. The prompting of these quicker responses would hence reduce overall variance in response times. The consequence of this would be the exacerbation of accuracy differences between blocks, given that more difficult trials tend to yield more erroneous responses when response times are constrained (Bolsinova, de Boeck, & Tijmstra, 2016). Notably, accuracy was less potent than response time in measuring relatedness in both experiments, as so facilitating a greater variance in the accuracy of responses in this way may prove beneficial at both the level of the cross-sectional approach to the FAST, as well as the meta-FAST. However, it is important to bear this conditional dependency in mind when implementing changes to the response window. Empirical analyses of differences resulting from such changes would be beneficial.

4.7 Summary and Conclusion

The current thesis intended to assess the effectiveness of a behaviour-analytic implicit measure, the Function Acquisition Speed Test (FAST), in quantifying differences in stimulus relatedness along inter-class and intra-class parameters. Establishing the FAST as an effective measure of stimulus relatedness would provide evidence for a behaviour-analytic account of implicit measures, as well as providing the field of stimulus equivalence research with a novel tool for indexing stimulus relations. As well as this, the current research intended to compare the typical method of scoring the FAST, the FAST score, with a novel scoring method developed by the author based on a Skinnerian conceptualisation of fluency (i.e., the rate-fluency differential; RFD score). Experiment 1 varied relatedness between arbitrary stimulus relations along inter-class parameters, with 4 different degrees of stimulus relatedness, varied using either a differing number of matching-to-sample (MTS) training iterations, or using Real Word stimuli of known relatedness (i.e., No Training, 1 Iteration, 3 Iterations, and or Real Word). Experiment 2 varied intra-class relatedness in a within-groups design, with subjects receiving MTS training in two 4-member stimulus classes and then completing three FASTs, each of which tested relations of differing nodal distance (A-B, A-C, and A-D; 0-node, 1-node, and 2-node, respectively). It was expected that increasing degrees of relatedness in both experiments would lead to commensurately larger differences between FAST blocks.

Both experiments conformed to what had been expected. That is, the FAST proved to be an effective measure of both inter-class and intra-class stimulus relatedness. In Experiment 1, RFD score was more sensitive than FAST score in quantifying differential relatedness, while in Experiment 2, FAST score was more sensitive than RFD score. Overall, the FAST's effectiveness in quantifying stimulus relatedness provides evidence for a behaviour-analytic, inductive account of implicit measures as quantifying differential stimulus relatedness. This account avoids the pitfalls of former behaviour-analytic and social-cognitive accounts of these effects, such as the 'REC model' and the 'implicit attitude' accounts. As well as this, these findings provide evidence for the FAST as a novel tool for measuring emergent stimulus relations with stimulus relations training procedures such as the MTS procedure. Although RFD score and FAST score proved differentially sensitive to differences in relatedness in the experiments, it is arguable that RFD is more conceptually-coherent and transparent than FAST score, as well as being more amenable to comparisons across slightly differing FAST procedures. A psychometric treatment of both measures is required, but the RFD score shows promise as an alternative measure to FAST score in quantifying effects.

As is the nature of inductive approaches to research, a great deal of new issues has emerged from the results of the current research. In particular, it would be beneficial for future research into stimulus relatedness to develop a new procedure which can more comprehensively control the stimulus control topographies of subjects in learning stimulus relations. The development of such a procedure would greatly aid future experimental research in elucidating the specific dynamics of effects in implicit measures. As well as this, a number of experiments are required in order to investigate whether inter-class and intra-class stimulus relatedness may interact in informing responses on the FAST. Beyond this, a goal for the FAST is its development into a procedure sensitive enough to measure within-subject differences in learning. However, barriers to this goal include potential issues regarding practice effects and FAST block order effects, and these barriers must be addressed before the FAST can be effectively adapted at the single-subject level. However, the FAST shows promise in this regard.

A final salient point for future development of the FAST is based on increasing its specificity in parsing differences between stimuli at higher degrees of relatedness. Efforts in this regard should be concentrated on increasing the difficulty of FAST blocks in order to increase the range of scores seen. In particular this is especially pertinent to the consistent block, where scores currently exhibit a ceiling effect after very little training. Further, the development of a meta-FAST procedure which incorporates temporal persistence of stimulus relations as a metric of relatedness would greatly increase the range of scores possible, and would likely be an effective index of relatedness (based on conceptual understandings of behavioural resistance and class compatibilities, as well as variance in the test-retest scores in the published IAT literature).

In conclusion, the current thesis has established that the FAST is effective in quantifying differences in terms of both inter-class and intra-class differential stimulus relatedness. This finding is of great importance in providing a stimulus relatedness account of implicit measures, as well as providing stimulus equivalence researchers with a novel tool for the measurement of the formation and strengthening of stimulus classes. While the development and research into this procedure is on-going, this thesis has provided the groundwork for the FAST as a behaviour-analytic, well-understood measure of differential stimulus relatedness.

References

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record*, *47*(2), 309-320. http://dx.doi.org/10.1007/bf03395227

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit Attitudes towards Homosexuality: Reliability, Validity, and Controllability of the IAT. *Experimental Psychology, 48*(2), 145-160. http://dx.doi.org/10.1026//0949-3946.48.2.145

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, r., & Stewart, I. (2006). Do You Really Know What You Believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a Direct Measure of Implicit Beliefs. *The Irish Psychologist*, *32*(7), 169-177.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A Sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) Model. *The Psychological Record*, *60*(3), 527-542. http://dx.doi.org/10.1007/bf03395726

Bentall, R., Jones, R., & Dickins, D. (1999). Errors and Response Latencies as a Function of Nodal Distance in 5-Member Equivalence Classes. *The Psychological Record*, *49*(1), 93-115. http://dx.doi.org/10.1007/bf03395309

Beums, S., Traets, F., De Houwer, J., & Beckers, T. (2017). Symmetry and Stimulus Class Formation in Humans: Control by Temporal Location in a Successive Matching Task. *Journal Of The Experimental Analysis Of Behavior*.

Binder, C. (1996). Behavioral Fluency: Evolution of a New Paradigm. *The Behavior Analyst, 19*(2), 163-197. http://dx.doi.org/10.1007/bf03393163

Bluemke, M., & Fiedler, K. (2009). Base rate effects on the IAT. *Consciousness And Cognition*, *18*(4), 1029-1038. http://dx.doi.org/10.1016/j.concog.2009.07.010

Bolsinova, M., & Maris, G. (2015). A test for conditional independence between response time and accuracy. *British Journal Of Mathematical And Statistical Psychology*, *69*(1), 62-79. http://dx.doi.org/10.1111/bmsp.12059

Bolsinova, M., de Boeck, P., & Tijmstra, J. (2016). Modelling Conditional Dependence Between Response Time and Accuracy. *Psychometrika*. http://dx.doi.org/10.1007/s11336-016-9537-6

Bolsinova, M., Tijmstra, J., Molenaar, D., & De Boeck, P. (2017). Conditional Dependence between Response Time and Accuracy: An Overview of its Possible Sources and Directions for Distinguishing between Them. *Frontiers In Psychology*, *8*. http://dx.doi.org/10.3389/fpsyg.2017.00202

Bortoloti, R., Rodrigues, N., Cortez, M., Pimentel, N., & de Rose, J. (2013). Overtraining increases the strength of equivalence relations. *Psychology & Neuroscience*, *6*(3), 357-364. http://dx.doi.org/10.3922/j.psns.2013.3.13

Calkin, A. (2005). Precision teaching: The Standard Celeration Charts. *The Behavior Analyst Today*, *6*(4), 207-215. http://dx.doi.org/10.1037/h0100073

Carr, D., & Felce, D. (2000). Application of stimulus equivalence to language intervention for individuals with severe linguistic disabilities. *Journal Of Intellectual And Developmental Disability*, *25*(3), 181-205. http://dx.doi.org/10.1080/13269780050144262

Cartwright, A. (2013). *Developing the Function Acquisition Speed Test for Homonegativity* (M.Sc). Maynooth University.

Cartwright, A., Roche, B., Gogarty, M., O'Reilly, A., & Stewart, I. (2016). Using a Modified Function Acquisition Speed Test (FAST) for Assessing Implicit Gender Stereotypes. *The Psychological Record*, *66*(2), 223-233. http://dx.doi.org/10.1007/s40732-016-0164-5

Cauraugh, J. (1990). Speed-Accuracy Tradeoff during Response Preparation. *Research Quarterly For Exercise And Sport*, *61*(4), 331-337. http://dx.doi.org/10.1080/02701367.1990.10607496

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26-58. http://dx.doi.org/10.2307/411334

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*(2), 257-271. http://dx.doi.org/10.3758/bf03204507

Colbert, D., Dobutowitsch, M., Roche, B., & Brophy, C. (2017). The proxy-measurement of intelligence quotients using a relational skills abilities index. *Learning And Individual Differences*, *57*, 114-122. http://dx.doi.org/10.1016/j.lindif.2017.03.010

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. http://dx.doi.org/10.1037/h0040957

Cullinan, V., Barnes-Holmes, D., & Smeets, P. (2000). A Precursor to the Relational Evaluation Procedure: Analyzing Stimulus Equivalence II. *The Psychological Record*, *50*(3), 467-492. http://dx.doi.org/10.1007/bf03395366

Cummins, J., Roche, B., Tyndall., I., & Cartwright, A. (In press). The relationship between differential stimulus relatedness and implicit test effect sizes. *Journal of the Experimental Analysis of Behavior.*

Dal Cin, S., Gibson, B., Zanna, M., Shumate, R., & Fong, G. (2007). Smoking in Movies, Implicit Associations of Smoking With the Self, and Intentions to Smoke. *Psychological Science*, *18*(7), 559-563. http://dx.doi.org/10.1111/j.1467-9280.2007.01939.x

De Houwer, J. (2008). Comparing measures of attitudes at the functional and procedural level: Analysis and implications. In R. Petty, R. Fazio & P. Brinol, *Attitudes: Insights from the new implicit measures* (pp. 361-390). New York: Psychology Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347-368. http://dx.doi.org/10.1037/a0014211

Devany, J., Hayes, S., & Nelson, R. (1986). Equivalence class formation in language-able and language-disabled children. *Journal Of The Experimental Analysis Of Behavior*, *46*(3), 243-257. http://dx.doi.org/10.1901/jeab.1986.46-243

Dickins, D., Bentall, R., & Smith, A. (1993). The Role of Individual Stimulus Names in The Emergence of Equivalence Relations: The Effects of Interpolated Paired-Associates Training of Discordant Associations Between Names. *The Psychological Record*, *43*(4), 713-724. http://dx.doi.org/10.1007/bf03395908

Dickins, T., & Dickins, D. (2001). Symbols, Stimulus Equivalence and the Origins of Language. *Behavior And Philosophy*, *29,* 221-244.

Dougher, M., & Markham, M. (1994). Stimulus equivalence, functional equivalence and the transfer of function. In S. Hayes, L. Hayes, M. Sato & K. Ono, *Behavior Analysis of Language and Cognition*(pp. 71-90). Reno, Nevada: Context Press.

Dougher, M., Augustson, E., Markham, M., Greenway, D., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal Of The Experimental Analysis Of Behavior*, *62*(3), 331-351. http://dx.doi.org/10.1901/jeab.1994.62-331

Doughty, A., Brierley, K., Eways, K., & Kastner, R. (2014). Effects of Stimulus Discriminability on Discrimination Acquisition and Stimulus-Equivalence Formation: Assessing the Utility of a Multiple Schedule. *The Psychological Record*, *64*(2), 287-300. http://dx.doi.org/10.1007/s40732-014-0001-7

Doughty, A., Leake, L., & Stoudemire, M. (2014). Failure to observe untested derived stimulus relations in extinction: Implications for understanding stimulus-equivalence formation. *Journal Of The Experimental Analysis Of Behavior*, *102*(3), 311-326. http://dx.doi.org/10.1002/jeab.111

Dube, W., Green, G., & Serna, R. (1993). Auditory successive conditional discrimination and auditory stimulus equivalence classes. *Journal Of The Experimental Analysis Of Behavior*, *59*(1), 103-114. http://dx.doi.org/10.1901/jeab.1993.59-103

Dymond, S., & Roche, B. (2013). *Advances in relational frame theory*. Oakland, CA: Context Press.

Egli, M., Joseph, B., & Thompson, T. (1997). Transfer of social attributions in stimulus equivalence classes by preschool children. *Psychological Reports*, *80*(1), 3. http://dx.doi.org/10.2466/pr0.80.1.3-21

Egloff, B., & Schmukle, S. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal Of Personality And Social Psychology*, *83*(6), 1441-1455. http://dx.doi.org/10.1037//0022-3514.83.6.1441

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and Unaided Response Control on the Implicit Association Tests. *Basic And Applied Social Psychology*, *27*(4), 307-316. http://dx.doi.org/10.1207/s15324834basp2704_3

Fields, L., & Moss, P. (2007). Stimulus relatedness in equivalence classes: Interaction of nodality and contingency. *European Journal Of Behavior Analysis*, *8*(2), 141-159.

Fields, L., & Watanabe-Rose, M. (2008). Nodal Structure and the Partitioning of Equivalence Classes. *Journal Of The Experimental Analysis Of Behavior*, *89*(3), 359-381. http://dx.doi.org/10.1901/jeab.2008-89-359

Fields, L., Arntzen, E., & Moksness, M. (2014). Stimulus Sorting: A Quick and Sensitive Index of Equivalence Class Formation. *The Psychological Record*, *64*(3), 487-498. http://dx.doi.org/10.1007/s40732-014-0034-y

Fields, L., Arntzen, E., Nartey, R., & Eilifsen, C. (2012). Effects of a Meaningful, a Discriminative, and a Meaningless Stimulus on Equivalence Class Formation. *Journal Of The Experimental Analysis Of Behavior*, *97*(2), 163-181. http://dx.doi.org/10.1901/jeab.2012.97-163

Fields, L., Landon-Jimenez, D., Buffington, D., & Adams, B. (1995). Maintained nodal-distance effects in equivalence classes. *Journal Of The Experimental Analysis Of Behavior*, *64*(2), 129-145. http://dx.doi.org/10.1901/jeab.1995.64-129

Franck, E., Raedt, R., & Houwer, J. (2007). Implicit but not explicit self-esteem predicts future depressive symptomatology. *Behaviour Research And Therapy*, *45*(10), 2448-2455. http://dx.doi.org/10.1016/j.brat.2007.01.008

Gatch, M., & Osborne, J. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal Of The Experimental Analysis Of Behavior*, *51*(3), 369-378. http://dx.doi.org/10.1901/jeab.1989.51-369

Gavin, A., Roche, B., & Ruiz, M. (2008). Competing Contingencies Over Derived Relational Responding: A Behavioral Model of the Implicit Association Test. *The Psychological Record*, *58*(3), 427-441. http://dx.doi.org/10.1007/bf03395627

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2014). *Bayesian data analysis*(3rd ed.). Florida: CRC Press.

Gray, N., Brown, A., MacCulloch, M., Smith, J., & Snowden, R. (2005). An Implicit Test of the Associations Between Children and Sex in Pedophiles. *Journal Of Abnormal Psychology*, *114*(2), 304-308. http://dx.doi.org/10.1037/0021-843x.114.2.304

Green, A., Carney, D., Pallin, D., Ngo, L., Raymond, K., Iezzoni, L., & Banaji, M. (2007). Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients. *Journal Of General Internal Medicine, 22*(9), 1231-1238. http://dx.doi.org/10.1007/s11606-007-0258-5

Greenwald, A., & Banaji, M. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27. http://dx.doi.org/10.1037//0033-295x.102.1.4

Greenwald, A., & Farnham, S. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal Of Personality And Social Psychology*, *79*(6), 1022-1038. http://dx.doi.org/10.1037//0022-3514.79.6.1022

Greenwald, A., & Nosek, B. (2008). Attitudinal dissociation: What does it mean?. In R. Petty, R. Fazio & P. Briñol, *Attitudes: Insights From The New Implicit Measures* (pp. 65-82). New Jersey: Lawrence Erlbaum.

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal Of Personality And Social Psychology, 74*(6), 1464-1480. http://dx.doi.org/10.1037//0022-3514.74.6.1464

Greenwald, A., Nosek, B., & Banaji, M. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal Of Personality And Social Psychology, 85*(2), 197-216. http://dx.doi.org/10.1037/0022-3514.85.2.197

Greenwald, A., Smith, C., Sriram, N., Bar-Anan, Y., & Nosek, B. (2009). Implicit Race Attitudes Predicted Vote in the 2008 U.S. Presidential Election. *Analyses Of Social Issues And Public Policy, 9*(1), 241-253. http://dx.doi.org/10.1111/j.1530-2415.2009.01195.x

Hall, G., Mitchell, C., Graham, S., & Lavis, Y. (2003). Acquired equivalence and distinctiveness in human discrimination learning: Evidence for associative mediation. *Journal Of Experimental Psychology: General*, *132*(2), 266-276. http://dx.doi.org/10.1037/0096-3445.132.2.266

Harrison, R., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal Of The Experimental Analysis Of Behavior, 54*(3), 225-237. http://dx.doi.org/10.1901/jeab.1990.54-225

Hausknecht, J., Halpert, J., Di Paolo, N., & Moriarty Gerrard, M. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal Of Applied Psychology*, *92*(2), 373-385. http://dx.doi.org/10.1037/0021-9010.92.2.373

Hayes, J. (2017). *Investigating the use of the Relational Evaluation Procedure (REP) in children*(Ph.D). National University of Ireland, Galway.

Hayes, S. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal Of The Experimental Analysis Of Behavior, 51*(3), 385-392. http://dx.doi.org/10.1901/jeab.1989.51-385

Hayes, S., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory* (1st ed.). New York: Kluwer Academic/Plenum Publishers.

Heathcote, A., Popiel, S., & Mewhort, D. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*(2), 340-347. http://dx.doi.org/10.1037//0033-2909.109.2.340

Hove, O. (2003). Differential Probability of Equivalence Class Formation Following a One-To-Many Versus a Many-To-One Training Structure. *The Psychological Record*, *53*(4), 617-634. http://dx.doi.org/10.1007/bf03395456

Hughes, S., Hussey, I., Corrigan, B., Jolie, K., Murphy, C., & Barnes-Holmes, D. (2016). Faking revisited: Exerting strategic control over performance on the Implicit Relational Assessment Procedure. *European Journal Of Social Psychology*, *46*(5), 632-648. http://dx.doi.org/10.1002/ejsp.2207

Hussey, I. (2017). *Open Source Implicit Relational Assessment Procedure*. Retrieved 28 September 2017, from https://osf.io/kg2q8/

Imam, A. (2001). Speed contingencies, number of stimulus presentations, and the nodality effect in equivalence class formation. *Journal Of The Experimental Analysis Of Behavior*, *76*(3), 265-288. http://dx.doi.org/10.1901/jeab.2001.76-265

Jajodia, A., & Earleywine, M. (2003). Measuring alcohol expectancies with the implicit association test. *Psychology Of Addictive Behaviors*, *17*(2), 126-133. http://dx.doi.org/10.1037/0893-164x.17.2.126

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal Of The Experimental Analysis Of Behavior*, *59*(2), 333-347. http://dx.doi.org/10.1901/jeab.1993.59-333

Knowles, E., Lowery, B., & Schaumberg, R. (2010). Racial prejudice predicts opposition to Obama and his health care reform plan. *Journal Of Experimental Social Psychology*, *46*(2), 420-423. http://dx.doi.org/10.1016/j.jesp.2009.10.011

Lakens, D., Schneider, I., Jostmann, N., & Schubert, T. (2011). Telling Things Apart. *Psychological Science*, *22*(7), 887-890. http://dx.doi.org/10.1177/0956797611412391

Lee, D. (2006). Facilitating Transitions Between and Within Academic Tasks. *Remedial And Special Education*, *27*(5), 312-317. http://dx.doi.org/10.1177/07419325060270050601

Lee, M., & Wagenmakers, E. (2013). *Bayesian cognitive modeling*. Cambridge: Cambridge University Press.

Lee, M., & Wagenmakers, E. (2013). *Bayesian cognitive modeling*. London: Cambridge University Press.

MacCorquodale, K., & Meehl, P. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*(2), 95-107. http://dx.doi.org/10.1037/h0056029

Mace, F., Lalli, J., Shea, M., Lalli, E., West, B., Roberts, M., & Nevin, J. (1990). The momentum of human behavior in a natural setting. *Journal Of The Experimental Analysis Of Behavior*, *54*(3), 163-172. http://dx.doi.org/10.1901/jeab.1990.54-163

Martens, B., Lochner, D., & Kelly, S. (1992). The effects of variable-interval reinforcement on academic engagement: a demonstration of matching theory. *Journal Of Applied Behavior Analysis*, *25*(1), 143-151. http://dx.doi.org/10.1901/jaba.1992.25-143

Mauss, I., Evers, C., Wilhelm, F., & Gross, J. (2006). How to Bite Your Tongue Without Blowing Your Top: Implicit Evaluation of Emotion Regulation Predicts Affective Responding to Anger Provocation. *Personality And Social Psychology Bulletin*, *32*(5), 589-602. http://dx.doi.org/10.1177/0146167205283841

McConnell, A., & Leibold, J. (2001). Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal Of Experimental Social Psychology*, *37*(5), 435-442. http://dx.doi.org/10.1006/jesp.2000.1470

McIlvane, W., & Dube, W. (1992). Stimulus Control Shaping and Stimulus Control Topographies. *The Behavior Analyst*, *15*(1), 89-94. http://dx.doi.org/10.1007/bf03392591

McIlvane, W., & Dube, W. (2003). Stimulus control topography coherence theory: Foundations and extensions. *The Behavior Analyst*, *26*(2), 195-213. http://dx.doi.org/10.1007/bf03392076

McKenna, I., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the Fake-ability of the Implicit Relational Assessment Procedure (IRAP): The First Study. *International Journal Of Psychology And Psychological Therapy*, *7*(2), 128-138.

Meehl, P. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy Of Science*, *34*(2), 103-115.

Molesworth, B., & Chang, B. (2009). Predicting Pilots' Risk-Taking Behavior Through an Implicit Association Test. *Human Factors: The Journal Of The Human Factors And Ergonomics Society*, *51*(6), 845-857. http://dx.doi.org/10.1177/0018720809357756

Moorman, R., & Podsakoff, P. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal Of Occupational And Organizational Psychology*, *65*(2), 131-149. http://dx.doi.org/10.1111/j.2044-8325.1992.tb00490.x

Morey, R., Romeijn, J., & Rouder, J. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *Journal Of Mathematical Psychology*, *72*, 6-18. http://dx.doi.org/10.1016/j.jmp.2015.11.001

Moss-Lourenco, P., & Fields, L. (2011). Nodal Structure and Stimulus Relatedness in Equivalence Classes: Post-Class Formation Preference Tests. *Journal Of The Experimental Analysis Of Behavior*, *95*(3), 343-368. http://dx.doi.org/10.1901/jeab.2011.95-343

Nederkoorn, C., Houben, K., Hofmann, W., Roefs, A., & Jansen, A. (2010). Control yourself or just eat what you like? Weight gain over a year is predicted by an interactive effect of response inhibition and implicit preference for snack foods. *Health Psychology*, *29*(4), 389-393. http://dx.doi.org/10.1037/a0019921

Nelson, D., McEvoy, C., & Schreiber, T. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved 9 October 2017, from http://w3.usf.edu/FreeAssociation/

Nevin, J. (1988). Behavioral momentum and the partial reinforcement effect. *Psychological Bulletin*, *103*(1), 44-56. http://dx.doi.org/10.1037//0033-2909.103.1.44

Nevin, J. (1992). Behavioral contrast and behavioral momentum. *Journal Of Experimental Psychology: Animal Behavior Processes*, *18*(2), 126-133. http://dx.doi.org/10.1037//0097-7403.18.2.126

Nevin, J., & Grace, R. (2000). Preference and resistance to change with constant-duration schedule components. *Journal Of The Experimental Analysis Of Behavior*, *74*(1), 79-100. http://dx.doi.org/10.1901/jeab.2000.74-79

Nock, M., & Banaji, M. (2007). Assessment of Self-Injurious Thoughts Using a Behavioral Test. *American Journal Of Psychiatry*, *164*(5), 820. http://dx.doi.org/10.1176/appi.ajp.164.5.820

Nock, M., Park, J., Finn, C., Deliberto, T., Dour, H., & Banaji, M. (2010). Measuring the Suicidal Mind. *Psychological Science*, *21*(4), 511-517. http://dx.doi.org/10.1177/0956797610364762

Nosek, B., Greenwald, A., & Banaji, M. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. Bargh, *Automatic Processes in Social Thinking and Behavior* (pp. 265-292). New York: Psychology Press.

O’Reilly, A., Roche, B., Ruiz, M., Tyndall, I., & Gavin, A. (2012). The Function Acquisition Speed Test (FAST): A Behavior Analytic Implicit Test for Assessing Stimulus

Relations. *The Psychological Record, 62*(3), 507-528. http://dx.doi.org/10.1007/bf03395817

O'Reilly, A., Roche, B., & Cartwright, A. (2015). Function over form: A behavioural approach to implicit attitudes. In Z. Jin, *Exploring Implicit Cognition: Learning, Memory, and Social Cognitive Processes* (pp. 162-182). Philadelphia: Hershey.

O'Reilly, A., Roche, B., Gavin, A., Ruiz, M., Ryan, A., & Campion, G. (2013). A Function Acquisition Speed Test for Equivalence Relations (FASTER). *The Psychological Record*, *63*(4), 707-724. http://dx.doi.org/10.11133/j.tpr.2013.63.4.001

Osborne, J., & Blanchard, M. (2011). Random Responding from Participants is a Threat to the Validity of Social Science Research Results. *Frontiers In Psychology*, *1*. http://dx.doi.org/10.3389/fpsyg.2010.00220

Pashler, H., & Baylis, G. (1991a). Procedural learning: I. Locus of practice effects in speeded choice tasks. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *17*(1), 20-32. http://dx.doi.org/10.1037//0278-7393.17.1.20

Pashler, H., & Baylis, G. (1991b). Procedural learning: II. Intertrial repetition effects in speeded-choice tasks. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *17*(1), 33-48. http://dx.doi.org/10.1037//0278-7393.17.1.33

Pierce, W., & Cheney, C. (2013). *Behavior analysis and learning* (5th ed.). New York: Psychology Press.

Pilgrim, C., & Galizio, A. (2000). Stimulus equivalence and units of analysis. In J. Leslie & D. Blackman, *Experimental and Applied Analysis of Human Behavior* (pp. 111-126). Reno, Nevada: Context Press.

Pilgrim, C., & Galizio, M. (1990). Relations between baseline contingencies and equivalence probe performances. *Journal Of The Experimental Analysis Of Behavior*, *54*(3), 213-224. http://dx.doi.org/10.1901/jeab.1990.54-213

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal Of The Experimental Analysis Of Behavior*, *63*(3), 239-254. http://dx.doi.org/10.1901/jeab.1995.63-239

Popper, K. (1934). *The logic of scientific discovery* (1st ed.). London: Routledge.

Pritchard, D., Hoerger, M., & Mace, F. (2014). Treatment relapse and behavioral momentum theory. *Journal Of Applied Behavior Analysis*, *47*(4), 814-833. http://dx.doi.org/10.1002/jaba.163

Radley, K., & Dart, E. (2016). Antecedent Strategies to Promote Children's and Adolescents' Compliance with Adult Requests: A Review of the Literature. *Clinical Child And Family Psychology Review*, *19*(1), 39-54. http://dx.doi.org/10.1007/s10567-015-0197-3

Rehfeldt, R., & Dymond, S. (2005). The Effects of Test Order and Nodal Distance on the Emergence and Stability of Derived Discriminative Stimulus Functions. *The Psychological Record*, *55*(2), 179-196. http://dx.doi.org/10.1007/bf03395505

Richetin, J., Costantini, G., Perugini, M., & Schönbrodt, F. (2015). Should We Stop Looking for a Better Scoring Algorithm for Handling Implicit Association Test Data? Test of the Role of Errors, Extreme Latencies Treatment, Scoring Formula, and Practice Trials on Reliability and Validity. *PLOS ONE, 10*(6), e0129601. http://dx.doi.org/10.1371/journal.pone.0129601

Richetin, J., Perugini, M., Prestwich, A., & O'Gorman, R. (2007). The IAT as a predictor of food choice: The case of fruits versus snacks. *International Journal Of Psychology, 42*(3), 166-173. http://dx.doi.org/10.1080/00207590601067078

Ridgeway, I., Roche, B., Gavin, A., & Ruiz, M. (2010). Establishing and Eliminating Implicit Association Test Effects in the Laboratory: Extending the Behavior-Analytic Model of the IAT. *European Journal Of Behavior Analysis*, *11*(2), 133-150. http://dx.doi.org/10.1080/15021149.2010.11434339

Roche, B., Ruiz, M., O'Riordan, M., & Hand, K. (2005). A relational frame approach to the psychological assessment of sex offenders. In e. Quayle & m. Taylor, *Viewing Child Pornography on the Internet: Understanding the Offence, Managing the Offender, and Helping the Victims* (pp. 109-125). Dorset: Russell House Publishing.

Röhner, J., Schröder-Abé, M., & Schütz, A. (2011). Exaggeration is Harder Than Understatement, but Practice Makes Perfect!. *Experimental Psychology*, *58*(6), 464-472. http://dx.doi.org/10.1027/1618-3169/a000114

Rothermund, K., & Wentura, D. (2004). Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations. *Journal Of Experimental Psychology: General*, *133*(2), 139-165. http://dx.doi.org/10.1037/0096-3445.133.2.139

Schultz, P., Shriver, C., Tabanico, J., & Khazian, A. (2004). Implicit connections with nature. *Journal Of Environmental Psychology*, *24*(1), 31-42. http://dx.doi.org/10.1016/s0272-4944(03)00022-7

Shahan, T. (2017). Moving Beyond Reinforcement and Response Strength. *The Behavior Analyst*, *40*(1), 107-121. http://dx.doi.org/10.1007/s40614-017-0092-y

Sidman, M. (1960). *Tactics of scientific research*. New York: Authors Cooperative, Inc.

Sidman, M. (2009). Equivalence Relations and Behavior: An Introductory Tutorial. *The Analysis Of Verbal Behavior*, *25*(1), 5-17. http://dx.doi.org/10.1007/bf03393066

Skinner, B. (1937). *The behavior of organisms*. Acton, Mass.: Copley Pub. Group.

Skinner, B. (1957). *Verbal behavior* (1st ed.). Vermont: Echo Point Books.

Skinner, B. (1959). *Cumulative record*. Acton, Massachsetts: Copley Publishing Group.

Skinner, B. (1971). *Beyond Freedom and Dignity*. Indianapolis: Hackett Pub. Co.

Smeets, P., Barnes, D., & Roche, B. (1997). Functional Equivalence in Children: Derived Stimulus–Response and Stimulus–Stimulus Relations. *Journal Of Experimental Child Psychology, 66*(1), 1-17. http://dx.doi.org/10.1006/jecp.1997.2378

Spencer, T., & Chase, P. (1996). Speed analyses of stimulus equivalence. *Journal Of The Experimental Analysis Of Behavior, 65*(3), 643-659. http://dx.doi.org/10.1901/jeab.1996.65-643

Spradlin, J., Cotter, V., & Braxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal Of Mental Deficiency, 5*, 556.

Steele, D., & Hayes, S. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal Of The Experimental Analysis Of Behavior*, *56*(3), 519-555. http://dx.doi.org/10.1901/jeab.1991.56-519

Steffens, M. (2004). Is the Implicit Association Test Immune to Faking?. *Experimental Psychology*, *51*(3), 165-179. http://dx.doi.org/10.1027/1618-3169.51.3.165

Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the Relational Evaluation Procedure. *The Psychological Record*, *54*(4), 531-552. http://dx.doi.org/10.1007/bf03395491

Stull, L., McGrew, J., Salyers, M., & Ashburn-Nardo, L. (2013). Implicit and Explicit Stigma of Mental Illness. *The Journal Of Nervous And Mental Disease*, *201*(12), 1072-1079. http://dx.doi.org/10.1097/nmd.0000000000000056

Teige-Mocigemba, S., Klauer, K., & Sherman, J. (2010). Practical guide to Implicit Association Task and related tasks. In B. Gawronski & B. Payne, *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (pp. 117-139). New York: Guilford Press.

Urcuioli, P., Wasserman, E., & Zentall, T. (2013). Associative Concept Learning in Animals: Issues and Opportunities. *Journal Of The Experimental Analysis Of Behavior*, *101*(1), 165-170. http://dx.doi.org/10.1002/jeab.62

Valentine, K., Buchanan, e., Scofield, J., & Beauchamp, M. (2017). *Beyond p-values: Utilizing Multiple Estimates to Evaluate Evidence*. *Open Science Framework*. Retrieved 23 September 2017, from https://osf.io/9hp7y

von Hippel, W., Brener, L., & von Hippel, C. (2008). Implicit Prejudice Toward Injecting Drug Users Predicts Intentions to Change Jobs Among Drug and Alcohol Nurses. *Psychological Science*, *19*(1), 7-11. http://dx.doi.org/10.1111/j.1467-9280.2008.02037.x

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social Categorization and Stimulus Equivalence. *The Psychological Record*, *41*(1), 33-50. http://dx.doi.org/10.1007/bf03395092

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, *58*(3), 475-482. http://dx.doi.org/10.1007/bf03395630

Appendix A

Consent Form for Conditions 1 thru 3 in Experiment 1:

In agreeing to participate in this research I understand the following:

This research is being conducted by Jamie Cummins (contact: jamie.cummins.2014@mumail.ie), a postgraduate student at the Department of Psychology, Maynooth University, under the supervision of Dr. Bryan Roche (contact: Bryan.T.Roche@nuim.ie / +353 (1) 708 6026). It is the responsibility of this student to adhere to professional ethical guidelines in their dealings with participants and the collection and handling of data. If I have any concerns about participation I understand that I may refuse to participate or withdraw at any stage.

I understand that this study involves examining the effectiveness of psychological computer-based implicit tests, which are purported to measure word association strengths. I will be exposed to a computer-based training procedure which involves learning to associate 'made-up' nonsense syllables, and these associations will then be tested on subsequent computer-based implicit test measures.

All data from the study will be treated confidentially and my data will not be identified by name at any stage of the data analysis or in the final report. The data will be compiled, analysed, and submitted in a postgraduate thesis. I understand that this data may also be used in as part of analyses for a scientific publication. All data collected will be retained in a locked filing cabinet or on a University computer in the department of psychology for the duration of the current research programme. No personally identifying information will be stored.

At the conclusion of my participation, any questions or concerns I have will be fully addressed.

I may withdraw from this study at any time, and may withdraw my data at the conclusion of my participation if I still have concerns.

I am over 18 years of age, and do not suffer from any condition which may make learning tasks stressful or harmful or if they are effected by any condition which may cause participation in computer-based tasks to be difficult.

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the National University of Ireland Maynooth Ethics Committee at **research.ethics@nuim.ie** or **+353 (0)1 708 6019**. Please be assured that your concerns will be dealt with in a sensitive manner.

Signed:

______________________Participant

______________________ Researcher

______________________ Date

Appendix B

Consent Form for Condition 4 in Experiment 1:

In agreeing to participate in this research I understand the following:

This research is being conducted by Jamie Cummins (contact: jamie.cummins.2014@mumail.ie), a postgraduate student at the Department of Psychology, Maynooth University, under the supervision of Dr. Bryan Roche (contact: Bryan.T.Roche@nuim.ie / +353 (1) 708 6026). It is the responsibility of this student to adhere to professional ethical guidelines in their dealings with participants and the collection and handling of data. If I have any concerns about participation I understand that I may refuse to participate or withdraw at any stage.

I understand that this study involves examining the effectiveness of psychological computer-based implicit tests, which are purported to measure word association strengths. I will be exposed to implicit tests which will be used to measure strengths of association between two 'real-life', innocuous word pairings.

All data from the study will be treated confidentially and my data will not be identified by name at any stage of the data analysis or in the final report. The data will be compiled, analysed, and submitted in a postgraduate thesis.. I understand that this data may also be used in as part of analyses for a scientific publication; however, I would not be identifiable at any stage through the dissemination of . All data collected will be retained in a locked filing cabinet or on a University computer in the Department of Psychology for the duration of the current research programme. No personally identifying information will be stored.

At the conclusion of my participation, any questions or concerns I have will be fully addressed.

I may withdraw from this study at any time, and may withdraw my data at the conclusion of my participation if I still have concerns.

I am over 18 years of age, and do not suffer from any condition which may make learning tasks stressful or harmful or if they are effected by any condition which may cause participation in computer-based tasks to be difficult.

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the National University of Ireland Maynooth Ethics Committee at **research.ethics@nuim.ie** or **+353 (0)1 708 6019**. Please be assured that your concerns will be dealt with in a sensitive manner.

Signed:

______________________Participant

_____________________ Researcher

_____________________ Date

Appendix C

The number of blocks of both training and testing completed by each participant who passed the MTS testing phase, in each iteration of the procedure. ‘S. 1’ refers to the first sitting of the MTS, etc.

One Iteration		S. 1		S.2		S. 3	
	Participant	Training	Testing	Training	Testing	Training	Testing
	1	4	3				
	2	2	1				
	8	5	3				
	9	4	1				
	14	4	2				
	16	4	2				
	24	2	1				
	27	6	3				
	35	5	3				
	36	3	2				

Three Iterations		S. 1		S.2		S. 3	
	Participant	Training	Testing	Training	Testing	Training	Testing
	6	5	4	1	1	1	1
	7	2	1	1	1	1	1
	12	2	1	1	1	1	1
	15	1	1	1	1	1	1
	25	2	1	1	1	1	1
	31	5	2	1	1	1	1
	32	4	1	1	1	1	1
	44	3	3	2	1	1	1
	45	3	2	1	1	1	1
	49	2	1	1	1	1	1

Appendix D

Consent Form Given to All Subjects in Experiment 2:

In agreeing to participate in this research I understand the following:

This research is being conducted by Jamie Cummins (contact: jamie.cummins.2014@mumail.ie), a postgraduate student at the Department of Psychology, Maynooth University, under the supervision of Dr. Bryan Roche (contact: Bryan.T.Roche@nuim.ie / +353 (1) 708 6026). It is the responsibility of this student to adhere to professional ethical guidelines in their dealings with participants and the collection and handling of data. If I have any concerns about participation I understand that I may refuse to participate or withdraw at any stage.

I understand that this study involves examining the effectiveness of psychological computer-based implicit tests, which are purported to measure word association strengths. I will be exposed to a computer-based training procedure which involves learning to associate 'made-up' nonsense syllables, and these associations will then be tested on a subsequent computer-based implicit test measure.

All data from the study will be treated confidentially and my data will not be identified by name at any stage of the data analysis or in the final report. The data will be compiled, analysed, and submitted in a postgraduate thesis. I understand that this data may also be used in as part of analyses for a scientific publication. All data collected will be retained in a locked filing cabinet or on a University computer in the department of psychology for the duration of the current research programme. No personally identifying information will be stored.

At the conclusion of my participation, any questions or concerns I have will be fully addressed.

I may withdraw from this study at any time, and may withdraw my data at the conclusion of my participation if I still have concerns.

I am over 18 years of age, and do not suffer from any condition which may make learning tasks stressful or harmful or if they are effected by any condition which may cause participation in computer-based tasks to be difficult.

If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the National University of Ireland Maynooth Ethics Committee at **research.ethics@nuim.ie** or **+353 (0)1 708 6019**. Please be assured that your concerns will be dealt with in a sensitive manner.

Signed:

______________________Participant

_____________________ Researcher

_____________________ Date

Appendix E

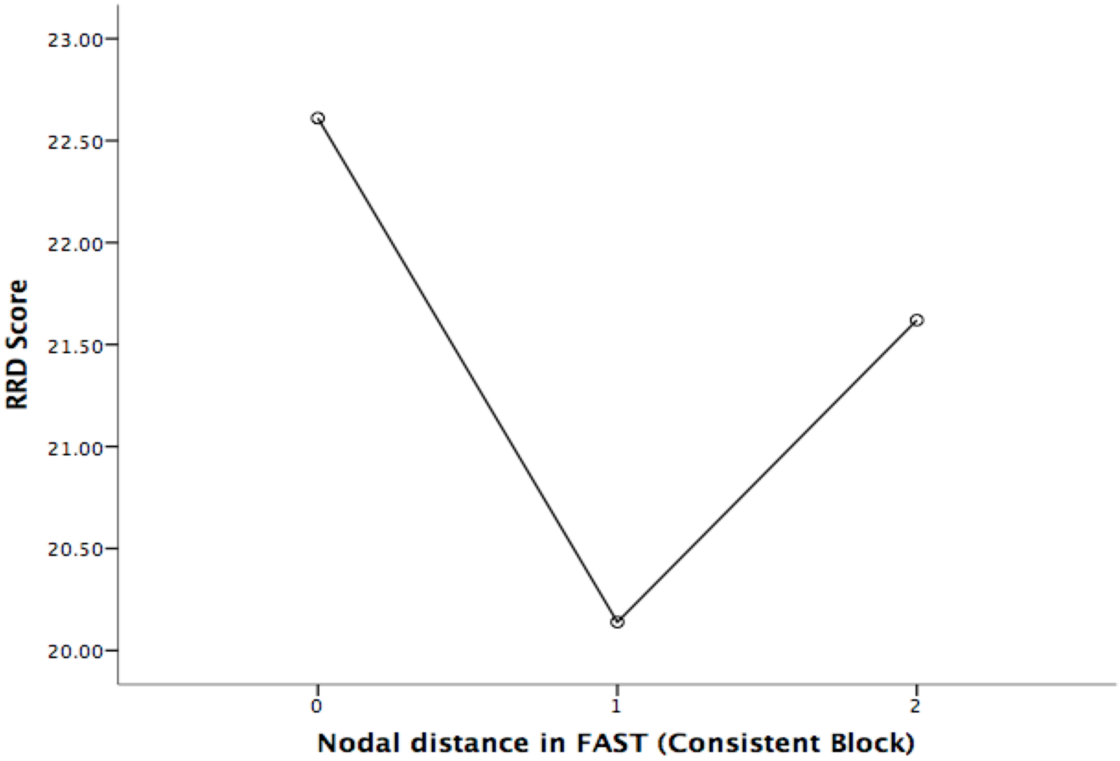
SPSS output for post-hoc analyses investigating the impact of differential nodal distance on scores on the consistent and inconsistent block of the FAST from Experiment 2.

Consistent block:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Consistent_Block	Sphericity Assumed	49.500	2	24.750	1.167	.325
	Greenhouse-Geisser	49.500	1.961	25.242	1.167	.324
	Huynh-Feldt	49.500	2.000	24.750	1.167	.325
	Lower-bound	49.500	1.000	49.500	1.167	.297
Error (Consistent_Block)	Sphericity Assumed	636.164	30	21.205		
	Greenhouse-Geisser	636.164	29.415	21.628		
	Huynh-Feldt	636.164	30.000	21.205		
	Lower-bound	636.164	15.000	42.411		



Inconsistent block:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Inconsistent_Block	Sphericity Assumed	150.919	2	75.459	3.574	.041
	Greenhouse-Geisser	150.919	1.892	79.783	3.574	.044
	Huynh-Feldt	150.919	2.000	75.459	3.574	.041
	Lower-bound	150.919	1.000	150.919	3.574	.078
Error (Inconsistent_Block)	Sphericity Assumed	633.470	30	21.116		
	Greenhouse-Geisser	633.470	28.374	22.325		
	Huynh-Feldt	633.470	30.000	21.116		
	Lower-bound	633.470	15.000	42.231