The relationship between differential stimulus relatedness and implicit measure effect sizes

Jamie Cummins[1], Bryan Roche[1], Ian Tyndall[2] & Aoife Cartwright[1]

[1] Maynooth University

[2] University of Chichester

Author note: The production of this manuscript was part-funded by a John and Pat Hume scholarship awarded to the first author for doctoral research conducted under the supervision of the second author. The software used in the current study is available by contacting the second author.

Abstract

Implicit measures have been hypothesized to allow researchers to ascertain the existence and strength of relations between stimuli, often in the context of research on attitudes. However, little ground-up and controlled behavioral research has focused on whether or not stimulus relations, and the degree of relatedness within such relations, are indexed by implicit measures. The current study examined this issue using a behavior-analytic implicit-style stimulus relation indexing procedure known as the Function Acquisition Speed Test (FAST). Using a matching-to-sample (MTS) procedure to train stimulus equivalence relations between nonsense syllables, the number of iterations of the procedure was varied across groups of subjects, hence controlling stimulus relatedness in the resulting equivalence relations. Following each subject's final exposure to the MTS procedure, they completed a FAST. An additional group of subjects was exposed to a FAST procedure with word pairs of known relatedness. Results showed that increasing relatedness resulted in a linear increase in FAST effect size. These results provide the first direct empirical support for a key process-level assumption of the implicit literature, and offer a behavior-analytic paradigm within which to understand these effects. These results also suggest that the FAST may be a viable procedure for the quantification of emergent stimulus relations in stimulus equivalence training.

Keywords: Function Acquisition Speed Test, stimulus equivalence, matching-to-sample, implicit measures, derived relational responding, Implicit Association Test.

The use of implicit psychological measures as a means of assessing the existence of stimulus relations or associations is common within social psychology research, and increasingly visible within experimental analysis of behavior (EAB) literature. Conceptions of these measures vary across paradigms. Social-cognitive theorists have argued that implicit measures are indicative of evaluative associations between mental representations, but have also been relatively conservative in attempting to precisely define evaluative associations (Greenwald & Nosek, 2008). EAB may offer a more fine-grained, empirically-based, functional account of implicit measure methodologies by appealing to well-understood behavioral processes, such as derived relational responding and stimulus class compatibilities (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010; Cartwright, Roche, Gogarty, O'Reilly, & Stewart, 2016; Gavin, Roche, & Ruiz, 2008). The present study highlights developments in one behavior-analytic test, the Function Acquisition Speed Test (FAST; O'Reilly, Roche, Ruiz, Tyndall, & Gavin, 2012), that aims to examine the degree of stimulus relatedness across classes of stimuli. The FAST could be considered a functional account of what social-cognitive theorists might refer to as an implicit attitude measure, built from the ground up using established behavioral principles. More specifically, the present study examines the utility of the FAST as a paradigm to potentially elucidate a key basic behavioral process account of whether or not stimulus relations, and the degree of relatedness within such relations, are actually indexed by implicit measures.

The best known implicit measure, known as the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), involves subjects responding rapidly on a computer keyboard based on rules that instruct a specific left-hand or right hand-response to each of four stimulus types (i.e., from one of four stimulus classes). Feedback is provided only for erroneous responses, with positive reinforcement never delivered. The rules are juxtaposed across two test blocks. For one block of trials the rule might state; "Press left for flowers and good words, press right for insects and bad words". For a second block, the rule might state; "Press left for flowers and bad words, press right for insects and good words". A difference score (*D*-score) based on the difference in response latencies divided by the pooled standard deviation of response times across blocks is used to infer the nature of the subject's attitudes regarding the tested stimuli. It is assumed that faster responses during one block compared to the other indicate associations (e.g., insects-bad / flowers-good) that are congruous with or informative of the subject's attitudes (see Greenwald & Banaji, 1995).

Implicit measures have also been developed within the behavior-analytic paradigm. One such measure is the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006). The IRAP originates from the theoretical position of Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), and possesses a number of procedural and scoring elements similar to the IAT, the main difference being that the IRAP allows for the specification of different types of relations between stimuli, whereas the IAT does not (see Barnes-Holmes et al., 2006, for a detailed overview of the procedure). Another behavioral measure, the FAST (O'Reilly et al., 2012), which is the focus of the current study, is more consciously directed towards the assessment of stimulus relations, both trained and derived, rather than primarily toward attitude measurement. The FAST is predicated on the principle that producing a shared functional response between two stimuli tends to be easier when the stimuli are members of the same stimulus class, and more difficult when stimulus classes differ (Roche, Ruiz, O'Riordan, & Hand, 2005). As the FAST was consciously built from the ground up using behavioral principles, one overriding goal was to avoid reliance on purely response time-based metrics as employed by the IAT and IRAP (i.e., the D-score method; see Ridgeway, Roche, Gavin, & Ruiz, 2010).

The current format of the FAST procedure (Cartwright et al., 2016) differs from the IAT in several ways (see Table 1 for an overview). One particularly salient aspect of the

FAST procedure is the presence of a response window. The IAT truncates response latencies over 3 seconds post hoc, whereas the FAST instead instantiates a 3 s response window. This response window influences the nature of responses registered in these procedures. Specifically, the presence of the response window punishes slow responding, and therefore results in generally quicker responding across both blocks. The prompting of these quicker responses hence reduces overall intra-subject variability in response times. 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). Conversely, the lengthening of response windows (such as in the IAT) ought to reduce intra-subject variability in accuracy of scores, while simultaneously exacerbating the effect of response time. Although this accuracy-response time conditional dependency is known, the precise weighting that a given response window places on response time or accuracy is currently unknown. In order to counteract this, the scoring of the FAST compounds response time and accuracy into a single metric (see Method).

[Table 1 around here]

The range of implicit measures both outside of and within behavior analysis (see Gawronski & DeHouwer, 2014; DeHouwer, Teige-Mocigemba, Spruyt, & Moors, 2009, for overviews) is indicative of differences in conceptual and procedural approaches to identifying and indexing stimulus relatedness. From a behavior-analytic perspective, stimulus relatedness is typically understood in terms of differential yield probabilities in stimulus discrimination testing, with more probabilistic yields indicating greater relatedness. In these paradigms, relatedness is typically varied in terms of nodal distance between stimuli in an equivalence class (e.g., Moss-Lourenco & Fields, 2011). However, relatedness has also been manipulated via overtraining of equivalence relations (Bortoloti, Rodrigues, Cortez, Pimentel, & de Rose, 2013). Despite the potential advantages offered by a behavioral approach in terms of procedural clarity and conceptual parsimony (Hughes, Barnes-Holmes, & DeHouwer, 2011), almost no published research has examined the efficacy of implicit measures in detecting laboratory-created and controlled stimulus relations, with the exception of two studies examining the FAST procedure (O'Reilly et al., 2012; O'Reilly et al., 2013) and two examining the IAT procedure, conducted in the same laboratory (Gavin, et al., 2008; Ridgeway et al., 2010; although see Hall, Mitchell, Graham, & Lavis, 2003, for a relevant study tapping into the same behavioral process harnessed by most implicit measures). The current study employed an approach similar to the foregoing studies, but also extended this experimental paradigm to incorporate the controlled variation of relatedness between experimental stimuli. To date, no study has systematically varied stimulus relations across experimental conditions and measured subsequent effects of this manipulation on implicit measures.

The current study was designed to answer two major questions. The first relates to the relationship between the size of effect in implicit measures and the relatedness of stimuli being examined. We will, for the purpose of considering this question, ignore the complicated issue of how response time–based scoring algorithms in the IAT and IRAP may render the behavioral processes underlying the basic effect rather opaque (see Gavin, Roche, Ruiz, Hogan, & O'Reilly, 2012; Ridgeway et al., 2010). While former studies have used a 'known-groups' approach to validate such measures (see Teige-Mocigemba, Klauer, & Sherman, 2010, for a discussion), such an approach does not provide a particularly reliable answer to this issue, because the variables against which implicit measure scores are correlated are themselves variable and uncontrolled. What is needed is an assessment of the variation in test effects as a function of experimentally-controlled or known (rather than inferred) degrees of stimulus relatedness.

The second question addressed by the current study relates to the potential utility of the FAST as a tool for the independent assessment of emerging derived stimulus relations (e.g., stimulus equivalence) at various stages of training. Specifically, stimulus equivalence relations are usually tested using accuracy criteria only (e.g., Fields, Adams, Verhave, & Newman, 1990), even though response time may also be of importance in determining how well a stimulus class has been established (Spencer & Chase, 1996; see also Fields, Arntzen, & Moksness, 2014, for an alternative assessment method and criteria). What may be of use, therefore, is an independent measure of relatedness that indexes both accuracy and response times of relational responding (see Spencer & Chase, 1996). A procedure that would allow for such indexing of degrees of stimulus relatedness would be of use in both laboratory research and in applied contexts. The current research, therefore, investigated the utility of the FAST procedure for this purpose.

In the current experiment, each subject was assigned to one of six conditions that differed in planned stimulus relatedness. One group received the FAST with real words of strong known relatedness, and the others received the FAST with arbitrary stimulus relations, the strength of which was manipulated via differential MTS iterations. Two of the conditions involved the same number of iterations of the MTS procedure, but the temporal spacing of training iterations was varied (i.e., the 3-in-1 condition completed 3 MTS procedures sequentially, while the 3-in-3 condition completed 3 MTS procedures separated by several days). This manipulation also allowed the authors to assess the FAST's sensitivity to differences in the relatedness of stimuli acquired using identical procedures, to identical criteria, but across varied temporal delays in training. Differentiating classes in this way is not a typical concern of EAB researchers, but such learning differences have been well-established in the associative learning field generally (see Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012, for a review), as well as specifically for conceptual stimulus categories (e.g., Vlach, Sandhofer, &

Kornell, 2008). 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, and in relation to the temporal distribution of training iterations. It was expected that responding to the FAST's consistent block would be more fluent with greater degrees of stimulus relatedness, while responding in the FAST's inconsistent block was expected to be less fluent with greater relatedness. In addition, it was expected that stimulus relations taken from the vernacular would produce the largest effect on the FAST (given that they were of the greatest relatedness due to subjects' pre-experimental histories).

Method

Subjects

The current experiment was approved for conduct by the University research ethics committee. The subjects within this study (*n* = 127) consisted of Caucasian, Irish undergraduate students attending Maynooth University and three other nearby universities. Subjects were recruited through the use of subject pools and sign-up sheets posted in the universities. Subjects received no remuneration for their participation. Subjects were quasi-randomly allocated to one of six conditions using a random number generator with a proportionate weighting of 2:1 for conditions that required MTS training compared to those that did not (No Training, *n* = 12; 1 MTS iteration [1 Iteration], *n* = 32; 2 MTS Iterations [2 Iterations], *n* = 25; 3 MTS iterations in one sitting [3-in-1], *n* = 20; 3 MTS iterations in 3 sittings [3-in-3], *n* = 21; and Real Words, *n* = 17). This proportionate weighting was introduced given the expected rate of attrition for subjects in the MTS. Seventy-three subjects identified as female, while the remaining 55 subjects identified as male. Thirty-eight subjects did not pass equivalence testing within four training and testing cycles (see Procedure), and hence were excluded from the study. The remaining 89 subjects had a mean age of 20.2 years (*SD* = 1.6 years), and consisted of 52 females and 37 males (No Training, *n* = 12; 1 Iteration, *n* = 23; 2 Iterations, *n* = 13; 3-in-1, *n* = 10; 3-in-3, *n* = 14; Real Words *n* = 17).

Apparatus

The experimental procedure was administered in a small, quiet research 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 room. All subjects engaged in all procedures on a 13” *Apple MacBook* with a screen resolution of 1024 x 768 pixels. The MTS training was 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.

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 coefficients which represent the probability that one stimulus will be discriminative of the response of another. For instance, the stimulus *cheddar* has a correlation coefficient of .922 with *cheese,* indicating 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* 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.

Procedure

Overview. Subjects were quasi-randomly allocated to one of the six experimental conditions. At the commencement of each experimental condition, subjects were asked to sign a consent document and were informed that they were free to withdraw from the experiment at any time. Subjects in the No Training and Real Word conditions completed a FAST with no training administered. The No Training condition's FAST involved nonsense stimuli, with which subjects had no prior history; by contrast, the FAST completed by subjects in the Real Word condition used real word stimuli of strong relatedness (e.g. salt-pepper). The four other conditions received a specific number of iterations of a MTS training and testing procedure designed to lead to the emergence of stimulus equivalence relations between two classes of nonsense syllables. If subjects did not reach criterion on both MTS training and testing in any one iteration within 40 minutes, the experimenter informed the subject that their participation was complete, debriefed them, and thanked them for taking part; this was the case for 38 subjects. Following completion of the MTS procedure, subjects in the 1 Iteration condition completed a FAST. Subjects in the 3-in-1 condition completed the MTS procedure twice more in the same sitting (with approximately 2-min intervals between commencement of each iteration of the procedure), and were then required to complete the FAST. Subjects in the 2 Iterations and 3-in-3 conditions were thanked for their participation, and returned to the testing booth approximately one week later for a second iteration of MTS training and testing. Subjects in the 2 Iterations condition completed the FAST after this second iteration. Subjects in the 3-in-3 condition completed one further MTS procedure a week following their second iteration, and were then administered the FAST immediately following this third MTS training and testing iteration. For all subjects, the FAST took no more than six minutes to complete. Stimulus relatedness was varied systematically via this stratification of conditions, in that it was assumed subjects who completed a greater amount of training would have a greater probability of responding in accordance with the trained relations.

Stimulus equivalence training and testing. The matching-to-sample procedure (MTS) was designed to establish two three-member equivalence relations among nonsense word stimuli. The stimuli CUG, JOM, VEK, LER, MAU, ZID were randomly assigned to stimulus classes (i.e., class 1 or 2) and roles (i.e., A, B, or C) within each class across subjects. The stimuli in the two predicted emergent classes will be referred to using the alphanumeric labels 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 center-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. Although it is typically preferred to use more than two comparison stimuli (see Carrigan & Sidman, 1992), the current study utilized only two stimulus classes (and hence, only two comparison stimuli) in order to limit the length of time required to complete the

MTS procedure, in line with ethical requirements of the university. There was a 1-s delay between sample and comparison presentations, which was intended to orient the subject towards the sample stimulus. 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 1 s. The feedback was accompanied by a brief auditory stimulus; a ‘beep’ sound for correct responses, or a lower-pitch ‘buzz’ sound for incorrect responses. An intertrial interval of 1 s, which consisted of a blank screen, followed the conclusion of the presented feedback.

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); responses which resulted in negative feedback are parenthesized. The four trial types were presented in a quasi-random order, with each trial presented once in a cycle of four trials. 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 of at least 31 correct responses in a block in order to progress to the equivalence test.

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 were present in this phase: A1-C1 (C2), A2-C2 (C1), C1-A1 (A2), and C2-A2 (A1). The four trial types were 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 32 trials per testing block.

No feedback was presented to subjects at any stage of the test phase. The criterion for passing this phase was also 31 correct out of 32 responses in 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. Training and testing cycles continued until the test was passed or had been administered up to a maximum of four times. Thirty eight subjects failed to pass the test after four testing cycles; when this occurred, subjects they were thanked for their time, debriefed, and their data were omitted from the study.

Function Acquisition Speed Test (FAST). Following the final exposure to the MTS procedure for those who received training, subjects who passed the testing phase were required to complete the 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 trials each. One block was designated the *consistent* block, and the other was designated the *inconsistent* block. The order of presentation of blocks was randomized across subjects by the FAST software (note that this randomization utilizes true randomness, i.e., there may not necessarily be an even number of subjects allocated into groups; see Table 2 for the specific distributions of this variable across groups). For subjects whose FASTs involved arbitrary stimuli, the consistent block involved the reinforcement of responses, in the presence of a given stimulus, which were 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). 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 the Real Word condition, the consistent block involved the reinforcement of responses congruous to those associations from the SFFANI (e.g., salt and pepper shared a response, and king and queen shared a response). The inconsistent block involved the opposite response patterns being reinforced (e.g., salt and king shared a response and pepper and 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.

[Figure 1 around here]

Following the first key press of the subject, the first intertrial interval (ITI) of 500 ms was presented (i.e., a blank white screen). Following this, a stimulus appeared in the center of the screen in size 32 point font, after which the subject was required to respond with the “Z” or “M” key. There was a 3000 ms time limit on responding, after which the corrective feedback for an erroneous response was presented. If a response was made within 3000 ms the screen was immediately cleared, and feedback was presented as appropriate. Both the consistent and inconsistent blocks of the FAST consisted of 50 trials; four different stimuli (A1, A2, C1, C2) were presented in conditions involving arbitrary stimuli, and two real-word pairs in the Real Word condition. The 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.

The scoring of the FAST procedure involves fitting a regression line to the cumulative learning curve (with cumulative response time on the X axis, and accuracy on the Y axis) generated by each FAST block (i.e., consistent and inconsistent; although see O’Reilly et al.

[2012, 2013] for an alternative scoring system based on a different FAST training and scoring format). The slope of this regression line is then calculated, and this value is referred to as the block-slope score. Higher block-slope scores are represented by a steeper regression line, which indicates a combination of less overall time taken (i.e., quicker response times) and more accurate responses in a given block. The block-slope score of the inconsistent block is subtracted from the block-slope of the consistent block. This value represents the difference between performances in the consistent and inconsistent blocks, and is referred to as the FAST score. A positive FAST score implies that responding was quicker and more accurate on the consistent block than the inconsistent block. A negative FAST score indicates the opposite; i.e., superiority of response performance on the inconsistent block relative to the consistent block.

Results

A total of 89 subjects completed the FAST with real word or with nonsense words following no training or the four MTS training conditions. Appendix 1 shows information on the number of blocks of training and testing completed by individual subjects who received MTS training. Subjects who completed more than one MTS iteration consistently reached criterion more quickly on subsequent iterations than on their first exposure to the procedure.

The primary measure of interest in the FAST was that of FAST score. Data were also collected on the number of accurate responses and mean response time of each block. Initial descriptive analyses showed that, as expected, a greater number of stimulus equivalence training iterations were associated with larger FAST scores. Table 3 displays the means of FAST scores for each of the six conditions, as well as the mean difference in total accurate responses, and difference in mean response times. Figure 2 illustrates the change in mean FAST scores across conditions, showing an increase in tandem with greater stimulus relatedness.

[Table 2 around here]

[Table 3 around here]

[Figure 2 around here]

Further descriptive analyses of performances within the FAST blocks revealed what appeared to be a general change in response time, accuracy, and block-slope on *both* consistent and inconsistent blocks across conditions. As Table 4 illustrates, mean block-slope scores and accuracy scores tended to be higher on the consistent block, and response times lower, as purported relatedness increased. By contrast, all three metrics showed a more mixed pattern of change on the inconsistent block.

[Table 4 around here]

Inferential analyses were used to test the effect of condition on FAST scores, total accuracy, and mean response time. These three metrics were normally distributed and did not violate assumptions of homogeneity of variance ($p = .36$, $p = .07$, and $p = .476$, respectively). Linear polynomial contrasts revealed a significant linear trend in FAST scores as a function of condition was (F (1, 83) = 14.414, $p = .0003$; $\eta_p^2 = 0.148$, observed power = .973), mean response time difference (F (1, 83) = 7.078, $p = .009$; $\eta_p^2 = 0.079$, observed power = .758), and mean accuracy difference (F (1, 83) = 14.949, $p = .0002$; $\eta_p^2 = 0.153$, observed power = .976).

A mixed-model 2 (FAST block) x 6 (training condition) ANOVA found a significant interaction effect between FAST block and training condition on block-slope scores, Wilks' Lambda = .887, F (5, 83) = 2.335, $p = .049$, $\eta_p^2 = .123$, observed power = .722 (see Figure 3). With increasing stimulus relatedness, block-slope scores for the consistent block generally increased, and plateaued from the 3-in-1 condition onwards. While block-slope scores for the inconsistent block also increased with initial increases in relatedness (though not to the same extent as the consistent block), the block-slope scores for the inconsistent block also began to decrease with increasing relatedness from the 3-in-1 condition onwards.

[Figure 3 around here]

Fifty-three subjects completed the consistent block first in the FAST, while 36 subjects completed the inconsistent block first (see Table 2). In order to investigate any potential confounding effect of block order on FAST scores, six independent t-tests were run and found no significant effect of block presentation order on FAST scores for any condition (No Training, t (10) = -.225, p = .826; 1 Iteration, t (21) = .725, p = .477; 2 Iterations, t (11) = .430, p = .676; 3-in-1, t (8) = .131, p = .57; 3-in-3, t (12) = .633, p = .392; Real Words, t (15) = .099, p = .856). In summary, all three metrics (FAST score, mean accuracy difference, and response time difference) fitted significantly to a linear trend as a function of experimental condition and there was a significant interaction effect between block and training condition for block-slope, with consistent block-slope scores rising as a function of training condition, while inconsistent block slope scores showed a more mixed trend.

Discussion

The current experiment demonstrated that increases in relatedness between stimuli resulted in increases in implicit measures effect sizes for those stimuli. Stimulus relatedness was varied systematically across six conditions, either through the number of iterations of the matching-to-sample (MTS) procedure, which involved the formation of two artificial relational classes (A1-B1-C1 and A2-B2-C2), or through the use of stimuli of known, strong relatedness. Stimulus relatedness was measured using the behavior-analytic FAST tool. The finding of a significant linear trend in FAST scores across conditions was consistent with the expected outcome that increases in relatedness between stimuli would lead to larger effects. Consistent with this also was a significant interaction effect between block and relatedness condition for block-slope scores on the 2x6 ANOVA. The findings from the trend analyses in particular suggest that greater relatedness resulted in linear increases in the difference between scores on consistent versus inconsistent blocks. This finding implies that effects on these measures may now be better understood as the consequence of stimulus relatedness in the individual’s learning history; that is, this experiment is the first of its kind to demonstrate that experimentally-controlled stimulus relatedness leads to subsequent changes in an implicit measure.

In addition, the data suggest that the FAST may have utility in measuring the differences 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 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 probe trials in a testing phase (as in the case of the MTS in the current experiment). 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). This indicates that the probe trial context and format may confound the phenomenon it intends to measure via directly taking part in the formation of equivalence relations. Card sorting tasks have been seen as a more sensitive alternative to probe trial formats (Fields, et al., 2014; Fields, Arntzen, Nartey, & Eilifsen, 2012), but answer only the discrete question of whether relations have formed or not.

The FAST offers a more continuous alternative; it avoids the issue of the probe-testing method given that stimuli are never matched in the procedure, provides more comprehensive data than card-sorting, and can be administered multiple times intermittently during equivalence training due to its short duration. Of course, multiple iterations of the FAST may potentially confound the training contingencies, and the FAST also does not provide a definitive outcome to determine whether or not equivalence relations have been formed. However, the FAST employs both equivalence 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. In addition, all test criteria for determining the formation of equivalence relations are ultimately arbitrary (e.g., 90% correct) and such criteria can easily be arrived at for the FAST method through further exploratory research. Notably, the current research (and previous research using the FAST with contrived equivalence relations, e.g., O'Reilly et al., 2013) involved the presentation of probe trials for equivalence classes prior to the administration of the FAST. Given that such probe trials may facilitate the formation of equivalence classes, it would be of interest to present the FAST to subjects in a similar paradigm to the current experiment immediately following MTS training, without intervening probe trials. Such a design would allow for a more controlled analysis of the FAST's utility in quantifying the emergence of derived relations. At this stage, however, the current findings suggest that the FAST has potential as a novel and robust measure of equivalence class formation and strength, which overcomes the issues of probe trial use, and is more data-enriched than card-sorting.

The current findings are also of interest in terms of how best to quantify the effects produced by implicit measures. In the analyses for FAST score, response time, and accuracy, the smallest effect size was for response time. This is of interest given that response time has been the conventional metric of interest in the field of implicit measures. The relative disparity in these effect sizes may be due to the issues that are associated with the typical method of parameterizing response times in this field. Specifically, the use of means and standard deviations to analyze response times 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). As such, it may be preferable in future studies using the FAST to examine response times through analyses along ex-Gaussian parameters. This could determine whether the differences in effect sizes are solely the consequence of ineffective parameterization of the response time data, and if not, to what greater degree FAST scores/accuracy difference scores are affected by differential stimulus relatedness relative to response time difference scores. Future studies may also seek to assess to what degree findings on other implicit measures are affected by this differential parameterization.

It is unknown whether further training on the arbitrary stimulus relations would have resulted in effects surpassing those seen in the Real Word condition, given that the 3-in-3 condition resulted in near-identical scores as the former. While block slope scores on the consistent block gradually increased, scores on the inconsistent block did not gradually decrease, and never dropped below the level of the No Training condition’s scores. Inconsistent block scores should likely decrease with increasing S- control, given that such cross-class matching should become more difficult. As such, S- control was likely not exerted to a great degree in any condition in the current experiment. Given that mastery in the MTS training phase subsequently reduces the number incorrect responses, which reduces opportunities to establish increasing S- control, it is likely that further MTS training may not result in greater effects. However, the use of some other training procedure which can control S- relations (such as an adapted version of the Relational Evaluation Procedure; see Stewart, Barnes-Holmes, & Roche, 2004) could hypothetically demonstrate greater effects with further iterations, as both S+ and S- control would increase over this time. Answering this question is salient in order to further understand the nuances of effects on implicit measures.

Further, in terms of MTS training and testing, it is notable that 38 subjects out of the original sample of 127 failed to reach the criterion of 31 out of 32 correct responses in the MTS testing within four cycles of testing. This rate of subject attrition is in at least some aspect attributable to the specific parameters of the training procedure employed. Specifically, many-to-one (MTO) and one-to-many (OTM) MTS protocols have been shown to produce higher yields in training stimulus equivalence relations than the linear MTS training utilized in the current study (Arntzen, 2012; Arntzen & Hansen, 2011). In addition, the current training employed a concurrent, rather than sequential, training structure, which has been shown to require more training trials than sequential training in the emergence of equivalence responding, as well as to produce less temporally-stable relations (Arntzen, 2012; Arntzen, Halstadtro, Bjerke, Wittner, & Kristiansen, 2014). The interspersing of baseline trials amongst equivalence probe trials during testing (rather than the exclusive presentation of probe trials only as employed in this study) has also been shown to facilitate yields in equivalence responding (Arntzen & Hansen, 2011; Tomanari, Sidman, Rubio & Dube, 2006). While the less-than-optimal MTS protocol employed in the current study does not detract from the veracity of the current findings (in the sense that all subjects who completed the FAST had achieved the criterion), future FAST research with a similar paradigm to the current work would benefit from the employment of a MTS protocol which is more effective in yielding equivalence responding.

The precise source of the effects seen in the current experiment is unclear, given that the number of both training and testing MTS trials varied across conditions. Future studies should seek to assess the effect of only administering training trials in a paradigm otherwise identical to the current. While this is of interest to investigate, unreported analyses for the current data revealed no significant correlation between the number of MTS training trials and FAST score, as well as no correlation between the number of MTS testing trials and FAST score. This would suggest that, in the current study, the number and temporal spacing of MTS iterations were the influential factor in manipulating relatedness, rather than sheer numbers of training and testing trials. This finding is coherent with previous research (e.g., Vlach et al., 2008), and suggests that future research using stimulus equivalence training and testing paradigms should consider temporal spacing as a significant variable in manipulating relatedness. Further, no other implicit measures were employed within this study, and hence it is unknown whether the current effect would be observed with other measures with procedural variants (e.g., absence of a response window). Future experiments should seek to execute the current paradigm with multiple implicit measures, in order to gauge the generalizability of the current findings. Any confluence of effects seen across test types would indicate similarity in the processes measured by these tests, while divergence would suggest that such tests may not be measuring learning histories per se.

A potential limitation of the current findings may be found in terms of the uneven distribution of block presentation order in the FAST across conditions. In particular, it is notable that 71% of subjects in the 3-in-3 condition were exposed to the consistent block of the FAST first. While this disparity in presentation order may suggest a confound, the data suggest that FAST scores increased regardless of which block was presented first (cf. Table 2). In addition, there was comparable similarity between scores within conditions regardless of presentation order, as indicated by the lack of significance on any of the 6 t-tests which were conducted. As such, it may be argued that block presentation order did not impact FAST scores, and therefore the experimental manipulation was the likely cause of the increases in FAST scores which were seen in the experiment.

A final area of inquiry may be centered on the further development of the sensitivity of these procedures. It is notable that there was no apparently large difference between FAST scores for stimuli trained in the 3-in-3 condition versus the Real Word condition, in spite of those stimuli in the Real Word condition being the subject of a much more enriched verbal history. It is arguable, therefore, that cross-sectional measures (e.g., a single FAST sitting) may not have the specificity to measure such differences in historical exposure to contingencies. The inclusion of an additional metric may provide greater precision in quantifying relatedness differences across stimulus sets. For example, temporal stability of a learned behavior can be an important indicator of the degree of class compatibility (Nevin & Grace, 2000). As such, a more effective means of quantifying relatedness could incorporate a measure of the rate of extinction of FAST scores across multiple FAST iterations, constituting what might be termed a “meta-FAST” procedure.

The current results demonstrated that FAST scores were affected by the degree of relatedness between stimuli. This research suggests that the FAST may also have utility as a means of assessing the emergence and strength of stimulus relations intermittently during training procedures, overcoming drawbacks which pertain to other attempts to do so. Building upon the current data, this research also has salient implications for furthering the debate regarding the appropriate metric of use in implicit studies. A novel metric exploring the persistence of FAST scores across FAST iterations and its relationship to stimulus relatedness may be of interest for future studies using the FAST, allowing for more comprehensive and diverse measures of relatedness compared to current cross-sectional approaches. In spite of these currently unexplored issues, the current research has demonstrated that FAST scores increase as a consequence of increasing stimulus relatedness, providing salient information for the behavior-analytic study of both stimulus equivalence and implicit measures.

Compliance with Ethical Standards

Declaration of Potential Conflicts of Interest: The authors declare they have no conflict of interest.

Research involving human subjects: All procedures performed in this study were in accordance with the ethical standards of the Maynooth University research ethics committee.

Informed Consent: Informed consent was obtained from all individual subjects included in the study.

References

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*(1). 123-135. http://dx.doi.org/10.1080/15021149.2012.11434412

Arntzen, E., Halstadtro, L., Bjerke, E., Wittner, K., & Kristiansen, A. (2014). On the sequential and concurrent presentation of trials establishing prerequisites for emergent relations. *The Behavior Analyst Today, 14*(1-2), 1-8. http://dx.doi.org/10.1037/h0101280

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal Of Behavior Analysis*, *12*(2), 483-503.

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.

Bolsinova, M., de Boeck, P., & Tijmstra, J. (2016). Modelling conditional dependence between response time and accuracy. *Psychometrika,* 1-23. https://doi.org/10.1007/s11336-016-9537-6.

Bortoloti, R., Rodrigues, N. C., Cortez, M. D., Pimentel, N., & de Rose, J. C. (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

Carpenter, S., Cepeda, N., Rohrer, D., Kang, S., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369-378. http://dx.doi.org/10.1007/s10648-012-9205-z

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58*, 183-204. http://dx.doi.org/10.1901/jeab.1992.58-183

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

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

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

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. H., Leake, L. W., & Stoudemire, M. L. (2014). Failure to observe untested derived stimulus relations in extinction: Implications for understanding stimulus-equivalence formation. *Journal of the Experimental Analysis of Behavior, 102,* 311-326. https://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

Fields, L., Adams, B., Verhave, T., & Newman, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of the Experimental Analysis of Behavior, 53*(3), 345-358. http://dx.doi.org/10.1901/jeab.1990.53-345

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

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. Available at: http://opensiuc.lib.siu.edu/tpr/vol58/iss3/6

Gavin, A., Roche, B., Ruiz, M., Hogan, M., & O'Reilly, A. (2012). A behavior-analytically modified Implicit Association Test for measuring the sexual categorization of children. *The Psychological Record, 62,* 55-68. http://doi.org/10.1007/BF03395786

Gawronski, B., & DeHouwer, J. (2014). Implicit measures in social and personality psychology. In: Reis, H. & Judd, C., *Handbook of Research Methods in Social and Personality Psychology* (2nd Ed., pp. 283-310). New York: Cambridge University Press.

Greenwald, A., & Banaji, M. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27.

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.

Greenwald, A., & Nosek, B. (2008). Attitudinal dissociation: What does it mean?. In R. Petty, R. Fazio & P. Brindol, *Attitudes: Insights from the New Implicit Measures* (1st ed., pp. 65-82). New Jersey: Erlbaum.

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,* 266-276.

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.

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

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109*(2), 340-347.

Hughes, S., Barnes-Holmes, D., & DeHouwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*, 465-496. Available at: http://opensiuc.lib.siu.edu/tpr/vol61/iss3/

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. https://dx.doi.org/10.1901%2Fjeab.2011.95-343

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/

Nevin, J., & Grace, R. (2000). Behavioral momentum: Empirical, theoretical, and metaphorical issues. *Behavioral and Brain Sciences, 23*(1), 117-125. https://doi.org/10.1017/S0140525X00502404

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://doi.org/10.11133/j.tpr.2013.63.4.001

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*, 507-528. http://doi.org/10.1007/BF03395817

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., & Galizio, M. (2000). Stimulus equivalence and units of analysis. In J. Leslie & D. Blackman, *Experimental and Applied Analysis of Human Behavio*r (1st ed., pp. 111-126). Nevada: Context Press.

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 and M. Taylor, *Viewing Child Pornography on the Internet: Understanding the Offence, Managing the Offender, and Helping the Victims* (1st ed., pp.109-125). Dorset: Russell House Publishing.

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

Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the relational evaluation procedure. *The Psychological Record, 54,* 531-552. http://doi.org/10.1007/BF03395491

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* (1st ed., pp. 117-139). New York: Guilford Press.

Vlach, H., Sandhofer, C., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163-167. https://doi.org/10.1016/j.cognition.2008.07.013

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475-482. Available at: http://opensiuc.lib.siu.edu/tpr/vol58/iss3/9

Table 1

An overview of some key differences between the IAT and FAST procedures.

	Procedure	
Feature	**IAT**	**FAST**
Purported to measure	Mental associations	Verbal learning histories (via stimulus class compatibilities)
Feedback	• Incorrect responses only • Subject must correct response • Error time-penalty also typically administered	All responses
Format	• 7 different blocks • 20 or 40 trials per block • 4th and 7th blocks are critical, others as 'practice'	• 2 blocks: one consistent, one inconsistent • 50 trials per block • All responses are critical
Response window	No response window in procedure, values are truncated or removed post-hoc	Active response window of 3000ms, failure to respond results in 'incorrect' feedback
Metric/scoring	D1 score: (Mean Incon RT – Mean Con RT) / Pooled RT_{SD}	FAST score: Difference in the slope of each blocks' learning curve

Table 2

Sample size, mean, and SD for FAST Scores across order of block presentation and experimental conditions.

Condition	Order	N (percentage)	Mean FAST Score	Std. Deviation
No Training	Con first	7 (58%)	-.0236	.07268
	Incon first	5 (42%)	-.0160	.02345
1 Iteration	Con first	13 (57%)	.0120	.06810
	Incon first	10 (43%)	-.0118	.08970
2 Iterations	Con first	7 (54%)	.0139	.04840
	Incon first	6 (46%)	-.0047	.10281
3-in-1	Con first	6 (60%)	.0400	.08064
	Incon first	4 (40%)	.0151	.02125
3-in-3	Con first	10 (71%)	.0392	.09883
	Incon first	4 (29%)	.0874	.06619
Real Words	Con first	10 (59%)	.0658	.12381
	Incon first	7 (41%)	.0563	.06417

Table 3

Mean differences in performance across FAST blocks for RT, accuracy, and FAST Score.

	Mean Response Time Difference (ms)	Mean Accuracy Difference (total correct)	FAST Score
No Training	-93.41 (208.05)	-1.33 (4.38)	-.02 (.06)
1 Iteration	-8.13 (325.63)	0 (4.6)	.00 (.08)
2 Iterations	44.71 (353.48)	.15 (5.55)	.01 (.08)
3-in-1	-8.9960 (199.46)	2.3 (3.74)	.03 (.06)
3-in-3	94.36 (223.23)	4 (6.92)	.05 (.09)
Real Words	207.66 (260.92)	4.82 (8.77)	.06 (.1)

Table 4

The mean FAST block-slope, accuracy, and response time scores across blocks and experimental conditions.

		Consistent	Inconsistent
Block-slope	No Training	.323 (.09)	.343 (.08)
	1 Session	.366 (.07)	.364 (.06)
	2 Sessions	.374 (.07)	.369 (.05)
	3-in-1	.433 (.03)	.403 (.06)
	3-in-3	.429 (.05)	.376 (.07)
	Real Words	.419 (.05)	.357 (.09)
Accuracy (Mean number of correct responses)	No Training	38.5 (7.19)	39.8 (6.09)
	1 Session	41.74 (3.9)	41.74 (4.23)
	2 Sessions	42.85 (5.23)	42.69 (3.09)
	3-in-1	45.6 (2.07)	43.3 (3.8)
	3-in-3	46.07 (2.73)	42.07 (6.11)
	Real Words	44.76 (3.23)	39.94 (7.77)
Response Time (ms)	No Training	1038.01 (323.4)	944.61 (232.58)
	1 Session	941.71 (289.56)	933.59 (201.22)
	2 Sessions	919.94 (277.53)	964.65 (223.82)
	3-in-1	780.89 (218.25)	771.89 (116.41)
	3 Sessions	777 (174.84)	871.36 (163.74)
	Real Words	685.28 (130.69)	892.94 (249.74)

Figure 1 The trained response functions for FAST stimuli across both the consistent and inconsistent blocks. Note that the consistent block involves stimuli from the same stimulus classes sharing functional responses, while the inconsistent block involves stimuli of different stimulus classes sharing responses.

Figure 2. Bar graph showing the change in mean FAST scores as a function of experimental condition. It can be seen that as the degree of relatedness between experimental stimulus classes increases, the mean FAST score of that experimental condition also increases.

Figure 3 Line graph showing the difference in mean block-slope scores in both consistent and inconsistent FAST blocks as a function of experimental condition. As degree of stimulus relatedness increases, scores on the consistent block generally increase, while scores on the inconsistent block initially increase, but then decrease.