MASTER'S THESIS
Learning in Complex Systems
Fall 2012

Speed and Accuracy in Fluency and Equivalence

Hastighet og nøyaktighet i etablering av flyt og ekvivalens

Pétur Ingi Pétursson

Institutt for Atferdsvitenskap

HØGSKOLEN I OSLO
OG AKERSHUS

Abstract

Speed and accuracy are the two main themes of the current thesis. Article 1 presents the conceptual, historical and empirical basis of behavioral fluency, how speed and accuracy are involved in fluency-based training and treatment, and the evidence for the superiority of fluency training. The possible implications of basic research on stimulus equivalence will be related to behavioral fluency, while considering a possible complementarity between the fields.

Article 2 is an empirical study on the comparison of speed and accuracy conditions in the establishment of stimulus equivalence and adduction and on the retention of the respective performances. There is some indication that a speed condition is most effective for the stability of equivalence relations, while equal probabilities of adduction result from either speed or accuracy conditions. The results are related to recent equivalence studies where speed and accuracy are involved.

Speed and Accuracy in Behavioral Fluency and Stimulus Equivalence: Implications from Applied and Basic Behavioral Research

Pétur Ingi Pétursson

Oslo- and Akershus University College

Abstract

Speed and accuracy are presented as major variables in a research endeavor known as behavioral fluency. Some of the historical origins of fluency research will be presented, with definitions of fluency, and a discussion of fluency training that aims at fluency as a behavioral outcome. Fluency training will be broken down into separate independent variables, and shown to collectively introduce a speed contingency on behavior, as related to a common fluency definition. Indicators of behavioral fluency will be discussed and whether current findings on fluency training support the explicit suggestions proposed by fluency theorists. Untrained, emergent performance has been postulated as one indicator of fluency, which makes another research endeavor known as stimulus equivalence, especially relevant since recently, analogous variables have been studied within that field. Empirical studies will be presented and discussed in relation to fluency theory and whether the results can be considered complimentary.

Speed and Accuracy in Behavioral Fluency and Stimulus Equivalence: Implications from Applied and Basic Behavioral Research

Behavioral fluency and precision teaching have gradually received increased attention in applied behavior analysis over the past 30 years. They are also popular in special education and increasingly used for therapeutic purposes, for example, for children with autism (Kubina & Wolfe, 2005; Kubina & Yurich, 2008). From 2001 to 2010, poster presentations on the ABAI on the use of fluency-based methods for this population rose from 12 to 175, while the number of empirical peer-reviewed articles published during the same period remained low (Heinicke, Carr, & LeBlanc, 2010). Some education literature, including, from its inception in 1980, the *Journal of Precision Teaching,* has published a number of empirical articles demonstrating the application of fluency training (Doughty, Chase, & O'Shields, 2004). In recent peer-reviewed demonstrations of its application, fluency training has been successfully applied in training programs for children with autism (e.g. Holding, Bray, & Kehle, 2011; Twarek, Cihon, & Eshleman, 2010), in the rehabilitation of children who have suffered traumatic brain injury (Chapman, Ewing, & Mozzoni, 2005; Kubina, Ward, & Mozzoni, 2000), to increase the math skills of pupils who are behind their class level (Chiesa & Robertson, 2000), to increase on-task endurance of students with ADHD (Brady & Kubina, 2010; McDowell & Keenan, 2001), and for enhancing children's reading (Downer, 2007), and spelling performance (Brooks, 1995).

Some studies comparing traditional education methods to precision teaching have shown that students are more likely to retain skills following a precision teaching than after a traditional approach. For example, Chiesa and Robertson (2000) showed that when five students who were behind in math received fluency training in component skills, they grew in scores from pre- to post-tests, as well as outperforming all of their classmates on composite math skills (see also Spangler & Hankins, 1975). However, empirical studies on fluency- vs.-

accuracy procedures in college teaching have yielded either higher post scores following an accuracy condition, or inconclusive results (Fox & Ghezzi, 2003; Péladeau, Forget, & Gagné, 2003).

Some examination of the history of the term fluency can help to explain the development of its current meaning and usage in the behavioral literature, and may shed some light on the reasons why fluency theorists have adapted the variables they use to influence behavior. Operational definitions of the treatment variables and behavioral processes or outcomes expected to encompass fluency will be identified. A number of effectiveness studies on fluency training exist, and may shed light on whether the desired behavioral outcomes are achieved with current fluency training technology. If not, then perhaps parallel variables studied in similar contingencies can provide partial solutions to unanswered research questions and contentions. This may be an example of a possible complementarity between applied and basic behavioral research.

Behavioral Fluency

In essence, three definitions of fluency are typically offered in the literature: one concerns its qualitative nature; the second, technical definition involves operations and measurement; while the third – the operational definition – bears more on the independent variables used for achieving fluency as an outcome. The qualitative behavioral definition shares the lay meaning of fluency (Binder, Haughton, & Bateman, 2002), and refers to qualitative features of behavioral repertoires, which are highly accurate, flowing, exerted with little effort and resistant to distraction. These can be either composite segments of complex repertoires or elementary component skills (Johnson & Layng, 1996). The second definition bears on the environmental arrangement or operations needed for achieving fluency: a continuous, free-operant performance, which means that an individual is free to respond to a stimulus or in the presence of a stimulus, without an imposed limitation on responding (Lindsley, 1996a).

The third definition appears to have determined how fluency is studied in empirical research, where speed has been selected as a major independent variable. Binder (1996) discussed some of the historical origins of fluency, the methods of behavioral fluency, and its shared history with precision teaching. Behavioral fluency can be traced back to the early 1960s, when Lindsley was starting to apply behavioral principles to treat people with developmental disabilities and to students' performance where rate was a main measure instead of accuracy (Binder, 1993). Precision teaching adopted the free-operant method. This meant that the environment was to be arranged so that no limits were imposed on responding, thus allowing influence and measurement of the temporal dimension. In the 1970s, Lindsley created the Standard Celeration Chart, which enabled both behavioral dimensions, accuracy and rate, to be shown on a single chart. This measurement system facilitated data sharing, so that data-based education decisions began to be made by teachers. This distinguished precision teaching from traditional teaching methods (Lindsley, 1996a). The Standard Celeration Chart is an important tool. Without a temporal dimension, an accuracy scale introduces a measurement ceiling on mastery (Binder, 2003).

Lindsley (1996a) discussed how the fluency concept was introduced into precision teaching. As timings of behavioral rate were applied to classroom behavior, the concept of behavioral fluency emerged. From 1967 to 1970, the forerunners of fluency training were using shorter and shorter timings to measure student classroom performance, reducing from about 40–50 minutes to 10 minutes. The reasons for this were practical: it allowed teachers to time a greater number of children each day. Seeing the benefits of using shorter timings, they soon moved to one-minute probes, to facilitate the measurement of a greater number of skill areas in a shorter time, pinpointing which area was in need of intervention. As this was occurring, Clay Starlin was researching interventions for dyslexic students, and got the idea that high-rate responding “would squeeze out errors and would permit generative curriculum leaps” (Lindsley, 1996b, p. 212). The timings moved from diagnostic-probe-based measurements to a specific intervention to increase skill rate in reading in one-minute timings, where speed was deemed important. In 1978, Lindsley took this idea further when he started a new training procedure in his university courses, which he called SAFMEDS (Say All Fast, a Minute, Every Day, Shuffle). He maintained that SAFMEDS would allow the students to achieve and maintain fluency through daily one-minute timing sessions (Potts, Eshleman, & Cooper, 1993). While fluency training often overlaps with precision teaching, Precision Teaching can better be described as a technology aimed at free-operant responding and measurement, as well as monitoring and sharing a student progress. Fluency training is a training component aimed at achieving fluency by increasing speed.

Speed and fluency training

Binder suggested “achieving true mastery adds speed to accuracy” (Binder, 2003, p. 15), and he has repeatedly defined fluency as accuracy and speed combined, both as an operation and as a process (Binder, 1993, 1996, 2010). In fact, the treatment variables used in fluency training probably introduce a speed contingency. Where empirical articles have provided concise definitions of treatment variables, these are mainly of three types:

(a) Rate aims: When responding becomes highly accurate, a higher rate is a usual goal of fluency training determined with individual aims for each skill (Binder, 2010). The practice involves presenting the aims and having the student try to reach the rate aims at their own pace in short timings of 30–60 s (Binder, 1996). In some cases the pace has been determined by the experimenter or therapist (e.g. Chapman et al., 2005; Kubina et al., 2000).

(b) Instructions: To reach the rate aims, students are either instructed or prompted to respond faster in timings occurring typically once a day, where rate and incorrect responses are charted (Merbitz, Vieitez, Merbitz, & Binder, 2004). For example, in a recent empirical study, participants exposed to a fluency training condition received the following instructions and prompts: “The participants were ‘coached’ to ‘go fast’ and reinforced for going quickly for a timed interval of 30 seconds” (Holding et al., 2011, p. 170; see also Brady & Kubina, 2010; Chapman et al., 2005; Chiesa & Robertson, 2000; Fox & Ghezzi, 2003; Kubina et al., 2000; Twarek et al., 2010).

(c) Reinforcement is made contingent upon the students increasing their rate on particular aims (Merbitz et al., 2004). In one study this was described as follows: “Once the fluency goal had been achieved, reinforcement was contingent on correct responses remaining at goal level” (Singer-Dudek & Greer, 2005 p. 180; see also Fox & Ghezzi, 2003; McDowell & Keenan, 2001).

To summarize, the key independent variables of fluency training are: (a) Rate or fluency aims; (b) differential reinforcement of a higher response rate; and (c) instructions or prompting to increase speed.

The significance of speed

The focus on speed is understandable given that fact that behavioral education often targets students who are behind in school and therefore need to be faster to catch up with their peers (Binder & Watkins, 1990). Response speed also reduces proportionally with age (Carriere, Cheyne, Solman, & Smilek, 2010), but can be increased with training independent of age level (Myerson, Robertson, & Hale, 2007). A certain pace is also important for certain types of skills such as motor, daily living, reading and writing skills (Twarek et al., 2010). The proportional devotion to speed in fluency training is however evident, independent of the type of skill targeted. Some fluency theorists have recommended that about 30% of instructional time be devoted to establish accurate responding and 70% to fluency training (Johnson & Layng, 1996). Lindsley recommended sometimes introducing speed before establishing high accuracy with the purpose of speeding up learning: “In building fluency… learners reach the instructional frequency aims sooner when they start out at high overall frequencies, showing steeper accelerations in frequency. They start at high speed… and low accuracy” (Lindsley, 1996a, p. 203). As fluency training is supposed to result in fluency as a behavioral outcome, dependent variables have been identified related to the qualitative definition of fluency.

Speed and fluency indicators

The three fluency indicators are retention, endurance and application performance standards, or REAPS (Dougherty & Johnston, 1996). Retention refers to the degree that performance is maintained over time, without direct training. Endurance is a measure of performance being maintained after a lapse of increasing intervals, and in the face of distractions. Application refers to component skills forming new composite skills in which the subject has never been directly trained (Binder, 1996). Adduction has been suggested as a fourth performance standard, where new behaviors are incorporated into established ones without further training (Johnson & Layng, 1996).

Fluency theorists have repeatedly stated that fluency training will lead to greater gains in terms of the performance standards, compared to an accuracy-only condition (e.g. Binder, 1993; Binder, 1996; Dougherty & Johnston, 1996). This has been the subject of a number of empirical demonstrations. For example, the retention of learned skills has repeatedly been shown to be higher following fluency training compared to accuracy-based or conventional methods (Berens, Boyce, Berens, & Kenzer, 2003; Bucklin, Dickinson, & Brethower, 2000; Fox & Ghezzi, 2003; Holding et al., 2011; Olander, Collins, McArthur, Watts, & McDade, 1986; Singer-Dudek & Greer, 2005; Spangler & Hankins, 1975), and some mixed results have also been reported (Brooks, 1995; Ivarie, 1986). A few studies have shown that fluency training results in a longer endurance than conventional or accuracy-based methods (Berens et al., 2003; Brady & Kubina, 2010; McDowell & Keenan, 2001), with a few studies having the same result with application (Berens et al., 2003; Bucklin et al., 2000). For example, Bucklin et al. (2000) evaluated the effects of accuracy or fluency training on both retention and application. Arbitrary stimuli were related through simple discrimination training and later tested in new combinations that the students had never experienced before. More untrained composite skills were shown to be retained following a fluency training condition after four and 16 weeks. In this instance however, the number of trials between conditions was not equal. No study known to this author has demonstrated adduction following fluency training. In fact, Singer-Dudek and Greer (2005) showed that composite skills did not emerge following fluency training of component skills, but that prompting was always needed.

The empirical status of behavioral fluency and performance standards has been a matter of considerable debate (see for example Binder, 2004; Chase, Doughty, & O’Shields, 2005; Doughty et al., 2004; Kubina, 2005). For example, Doughty et al. (2004) cited evidence presented by Lattal (1989) and Nevin, Grace, Holland and McLean (2001) which showed that with constant reinforcement rate, lower response rates are less affected by distraction. In a review paper by Doughty et al. they noted that between the years of 1983 and 2004 “only three studies yoked the number of practice trials between rate-building conditions and accuracy-only conditions and controlled for rate of reinforcement” (Doughty et al., 2004, p. 15). The three empirical studies (Evans & Evans, 1985; Evans, Merger, & Ben, 1983; Shirley & Pennypacker, 1994) were also reported to have produced mixed results. In addition none “tested for stability, application, or adduction while controlling for practice and rate of reinforcement” (Doughty et al., 2004, p. 17).

To exclude confounding when comparing fluency and accuracy conditions, an equal trial number must isolate response speed or inter-response time as an independent variable, given equal practice time. Yoking trial number and reinforcement rate allows the two variables to be examined simultaneously (Chase et al., 2005; Doughty et al., 2004). Of course each instructional trial is more than a simple stimulus exposure. It involves two discriminated operants in separate three-term contingencies, which become interlocked as one learning unit. In other words, it is an interaction between a student and an instructor or an instructional device (Greer & McDonough, 1999). Yoking learning units would allow speed to be isolated while holding both trial number and reinforcement constant.

Separating speed and accuracy

Since the review paper of Doughty et al., at least three articles have recently isolated speed and accuracy and thus controlled practice between training conditions. Singer-Duek and Greer (2005) used a simultaneous matched treatment design with four participants, matched on tests for prerequisite skills. Two of the participants received a fluency condition and two other participants an accuracy condition, yoked to the learning units from the fluency condition. For each condition, component tasks involved single-digit multiplication, addition and word reading. Following training, the participants were tested on component tasks they had not experienced before. Only one fluency participant reached criteria on the composite tasks earlier than the accuracy participant to whom he was yoked. However, on a two-month retention test on the composite skills, all the fluency participants reached criteria earlier than the accuracy participants. Contingency adduction from the trained skills did not result for any of the participants, as some prompting was necessary to reach performance criteria on the composite skills.

Holding et al. (2011) used a multiple treatment design with cross-subject yoking of discrete-trial training to fluency training and a variable reinforcement schedule being constant across participants. They found that retention of expressive noun skills established through a fluency condition was higher after six weeks than retention of those established through an accuracy condition. Fox and Ghezzi (2003) compared 18 college students who received fluency training with 18 college students who received accuracy training in concept analysis, randomly yoked to the trial number of another group of fluency participants, and already matched on pre-scores. Half of each group received either training in concept examples or in concept definitions. The results showed no differences between the groups that were either statistically significant or anything above marginal.

To summarize, most recent evidence suggests that where speed has been isolated with a yoking procedure, it has been shown to increase later retention, while limited empirical evidence exists for other indicators of fluency (see also Doughty et al., 2004).

Speed and emergence of untrained performance

Evidence that fluency training results in adduction could not be found. Contingency adduction involves novel repertoires emerging without shaping behavior, but instead recombining already established repertoires through reinforcement (Andronis, Layng, & Goldiamond, 1997). Examples of this include generalized imitation, as when children imitate actions they have never seen before, and generalized matching, as when people match stimuli that have previously been only indirectly related through training (Catania, 2007). This is one behavioral model used to account for what has been called discovery learning, in which skills emerge without training (Layng, Twyman, & Stikeleather, 2004).

Stimulus equivalence

Another behavioral model for emergent performance is stimulus equivalence. Matching-to-sample (MTS) procedures have been used to study stimulus equivalence where conditional discrimination is trained in discrete trials. Unlike the free-operant procedure, where one response occasions the next in a three-term contingency, in discrete trials only a single target response is available to a discriminative stimulus, which becomes reinforced and offsets another trial (Catania, 2007). The conditional discrimination is a four-term contingency (Sidman, 1994). With a MTS procedure, an organism is presented with a single conditional stimulus (S_{cond}), along with discriminative stimuli ($S^{D1,2,3\dots}$ and $S^{\Delta 1,2,3\dots}$). In the presence of one conditional (S_{cond1}) stimulus, responding (R) to a specific discriminative stimulus (S^{D1}) is reinforced (S^{r}). The experimenter has defined this as a correct selection. In effect, responding to the S^{Δ} s is not reinforced and is defined as an incorrect selection. In the presence of another conditional stimulus, S_{cond2}, responding to another discriminative stimulus S^{D2} is reinforced. Then responding to other discriminative stimuli is unreinforced (Sidman, 2000). An organism that responds according to the experimenter-defined conditionality is said to display conditional discrimination. Conditional discrimination performance resembles matching performance, but is not the same (Sidman, 1994).

In a typical experiment, the number of stimulus sets is specified by the number of alphabetical letters. Therefore ABCD refers to four sets of stimuli. One class is specified by a number that applies to members across sets, for example A1B1C1D1 refers to one class of stimuli with four members. A member of a class is specified by individual letter-number combinations, in effect A1 is one member and B1 is another (Green & Saunders, 1998). With B1 (S_{cond1}) presented as a sample, selecting A1 (S^{D1}) in the presence of A1 A2 A3 (S^{D1} and $S^{\Delta 2,3}$) represents a correct selection.

In a classical study, Sidman and Tailby (1982) trained eight participants to match four sets of arbitrary stimuli, A, B, C and D, and showed that after reaching experimenter-defined accuracy criteria, performance moved beyond the A-B, A-C and D-C trained conditionality, and untrained stimulus relations between members of sets B-A, C-A, C-D, B-C, C-B, A-D, B-D and D-B emerged. Arntzen (2012) described the relative amount of trained vs. emergent relations resulting from training potentially three 4-member classes:

> Trained relations may be expressed as C x (M-1), and emergent relations could be expressed as C x $(M-1)^2$, where C is the number of classes and M is the number of members. For example, if training potentially three 4-member classes – A1B1C1D1,

> A2B2C2D2, and A3B3C3D3 – there are nine trained relations (A1B1, A2B2, A3B3, B1C1, B2C2 B3C3, C1D1, C2D2, and C3D3) and 27 emergent relations (B1A1, B2A2, B3A3, C1B1, C2B2, C3B3, D1B1, D2B2, D3B3, A1C1, A2C2, A3C3, C1A1, C2A2, C3A3, B1D1, B2D2, B3D3, D1B1, D2B2, D3B3, A1D1, A2D2, A3D3, D1A1, D2A2, and D3A3 (p. 124).

The trained, together with the emergent relations, form an equivalent stimulus class (Sidman, 1994, p. 314). For a stimulus class to be equivalent, each relation has to show specific emergent properties. Mathematical equivalence is defined by reflexivity, symmetry and transitivity (Sidman, 2000). A behavioral specification of mathematical equivalence is made through behavioral tests which determine each property (Sidman, 1992). Reflexivity (if A-B then A-A and B-B), symmetry (if A-B, then B-A) and transitivity (if A-B and BC, then AC) can be independently demonstrated properties of responding, while a test for equivalence (if A-B, and BC, then CA) requires both properties (Sidman, 1992; Sidman & Tailby, 1982). The bi-directionality and transgression of function beyond the function found in simple and conditional discriminations shows how this is the most basic behavioral model for reference and meaning (Sidman, 1994). It demonstrates how skills emerge without requiring direct training (Arntzen, 2011). The amount of academic time, and time devoted to training or treatment, offers one suggestion of the applied value of stimulus equivalence. Bucklin and Dickson suggested that "stimulus equivalence tasks lend themselves well to studies of application because two unrelated component tasks are trained, and then individuals are asked to complete a third task that requires the component tasks but is not directly trained" (Bucklin et al., 2000, p. 148). Stimulus equivalence therefore offers a model for the investigation of parallels to application and adduction.

Complementary methodologies?

One challenge is that performance in discrete trials is limited by conditional discrimination and cannot result in free-operant performance (see Lindsley, 1996a). However, discrete trials may allow the analysis of moment-to-moment response changes that the free-operant procedure does not currently permit. For example, with the free-operant method speed is synonymous to rate; it is represented by a completion of some trial number within a particular duration. In another example, with a yoking procedure higher speed is represented by the lesser of two durations in an equal number of trials, which could be accounted for by higher response speeds in individual trials, or lower inter-response or inter-trial intervals. With a discrete trial method it is possible to keep these variables constant while measuring individual response latencies. An inverse of a response latency is a measure of the duration (speed) of each response per trial (Baron, 1985). The complementary information is the rate of responses vs. the speed of individual responses. This has, for example, made it possible to analyze a trade-off between accuracy and speed (see Imam, 2001). An important topic in stimulus equivalence research has been to identify the conditions for which stimulus equivalence is more or less likely to be established, and training conditions with a differential probability of equivalence outcomes (Arntzen, 2012). Among these are speed and accuracy, both as outcomes and as training conditions.

Research questions from fluency answered within an equivalence paradigm

Some parallel conditions and variables found in behavioral fluency can also be found in equivalence research. Based on response latency or speed, inferences have been made as to which variables influence equivalence formation, such as mediating behavior or structural features of equivalence classes (e.g. Arntzen & Lian, 2010; Holth & Arntzen, 2000; Spencer & Chase, 1996; Wulfert & Hayes, 1988). Some experimenters have attempted to reinforce lower latencies differentially in conditional discrimination training, to make inferences on the sources of equivalence classes (e.g. Arntzen & Haugland, 2012; Holth & Arntzen, 2000; Imam, 2001, 2003, 2006; Tomanari, Sidman, Rubio, & Dube, 2006). Such results should provide some answers to whether faster responding is more likely to result in stimulus equivalence, and therefore emergent performance. Once implicit contention by fluency theorists is that a speed contingency should be more likely to result in faster responding than trial repetitions alone (accuracy contingency). An additional contention, possible to answer with a discrete trial procedure, is whether it is possible to accelerate highly accurate performance, while holding the error of responding constant, so that high accuracy is not affected by increased speed (e.g. Binder, 2003; Lindsley, 1996a). Additional variables important to the study of equivalence possibly extend previous and current research on speed as a response dimension.

Speed and latency in equivalence: Further implications

Spencer and Chase (1996) demonstrated that when there is no difference in accuracy measures for directly trained and emergent relations, response speed can vary as a function of both relational types and nodal number. Relational types refer to unique combinations of trained or emergent stimulus relations (Imam, 2001), and a node refers to a stimulus connected to one or more stimuli through training (Fields & Verhave, 1987). Baseline and symmetry trials showed the fastest response speed on average, while transitivity and equivalence trials were the slowest. Response speed also varied as a function of nodal number, with a higher number of nodes resulting in slower response speeds (Spencer & Chase). This suggests that response speed can in some cases be expected to vary as a function of variables other than trial number or a direct manipulation of response speed. This may extend the previous findings in behavioral fluency. It suggests that a training design that systematically arranges for stimuli or training material to result in differential nodal numbers can be a determinant of the speed of highly accurate responding (i.e. fluency), without directly manipulating speed. In addition, some studies have shown that a training structure that results in classes with a higher nodal number, also results in lower response speed and accuracy to relations separated with a greater number of nodes (Bentall, Jones, & Dickins, 1998; Fields, Landon-Jimenez, Buffington, & Adams, 1995). Within an equivalence paradigm, fluency as a combination of accuracy and speed has to be partially a result of structural variables.

In the study by Chase and Spencer, equivalence was established in all subjects, probably because they used a one-to-many training (OTM) structure, which has a high probability of equivalence class formation. A linear-series (LS) structure has a low probability of equivalence class formation, as previous studies have revealed that about 10%–12% of subjects establish equivalence with this structure (Arntzen, Grondahl, & Eilifsen, 2010). Holth and Arntzen (2000) exposed 10 subjects to a training using an LS structure in establishing potentially three three-member classes. The result was that only one participant established an equivalence class. As in the Spencer and Chase study, symmetry and equivalence trial types in testing had a longer latency on average, which was unrelated to the class-consistency of responding. Of particular interest, however, was whether response latency would systematically vary as a function of class consistency in responding, or of consistency that was not related to stimulus equivalence. Response latency to comparison stimuli turned out to be lower in training and testing with class-consistent responding. For the purpose of analyzing the type of responding that may precede equivalence class formation, they manipulated a limited hold of the comparison stimuli down to 2s. The 10 subjects were trained with a one-to-many (OTM) structure, which has a higher probability of equivalence formation than an LS-structure (Arntzen et al., 2010). This led to lower response latencies, but the faster responding resulted in an immediate accuracy decline in tests for emergent relations, and equivalence was not established with any subject. Although accurate responses are generally faster, it seems that a speed condition may decrease emergent performance.

Imam (2001) noted that the response speed reduction reported by Spencer and Chase could have been accounted for by an unequal number of trials devoted to training the later stimulus relations (e.g. AD – three nodes) compared to the previous ones (e.g. AB – no node). In Experiment 1 he replicated the finding of Spencer and Chase. In Experiment 2 he made an adjustment so that an equal number of trials were devoted to training each stimulus relation. He found that that response accuracy and speed remained unchanged with different nodal numbers, suggesting that the effect might be an artifact of an unbalanced trial number in training and testing. As was previously reported, however, response speed tended to decline as a function of relational type, rather than nodal number. With a speed contingency, both response speed and accuracy varied as a function of relational type and nodal number. Without a speed contingency, response accuracy did not vary as a function of nodal number or relational type, but response speed varied as a function of both when accuracy had peaked (Imam, 2001). Similar results have been replicated once (see Imam, 2006). In both studies error rates were generally higher in baseline and tests for emergent relations in an accuracy and speed compared to an accuracy-only contingency, again showing that speed reduced the probabilities for stimulus equivalence.

What limits this comparison to fluency research is the fact that in these studies the speed contingency was not superimposed onto highly accurate responding, but at the start of baseline training. In addition, the three previously mentioned studies used different stimulus sets for the speed and accuracy conditions. In fluency training, speed is always introduced to an already mastered set of training material. The comparison would be more relevant if response speed were influenced by the same stimulus set in training blocks that already had high accuracy.

In two recent equivalence studies a limited hold to comparison stimuli was gradually reduced, so that if criterion-accuracy was reached in individual blocks, the limited hold was reduced in a stepwise manner. The resemblance to behavioral fluency methods is that, unlike in the previous studies, speed was introduced to the same stimulus set and increased gradually only after high accuracy had been established in individual blocks. Tomanari, Sidman, Rubio and Dube (2006) trained five participants in conditional discrimination of three sets of potentially four-member classes, using an OTM training structure. During training, a limited hold was gradually lowered from 0.4–0.5 s with the sample stimuli, and from 1.2–1.3 s with the comparison stimuli, as well as using an intertrial interval (ITI) of 0.4 s. Under these conditions the participants responded fast, and three of the five participants responded according to equivalence. The fact that only three of five participants responded according to equivalence is contrary to the prediction previously offered, in that untrained performance was expected to be greater in conditions that resemble fluency training. What limits this comparison is that fluency training sessions are usually much shorter in duration, and the possible effect of fatigue in a speed contingency could influence the probability of equivalence.

Arntzen and Haugland (2012) performed a systematic replication of the Tomanari et al. study with the modification that only comparison stimuli were presented with a limited hold to five experimental participants. As previously, the limited hold decreased as a function of high accuracy, in this case down to 1.2 s. In addition, the training sessions were much shorter at about 30 minutes. This did however reduce the probability of stimulus equivalence, as only one of five participants finally responded according to equivalence, and the fact that speed was much lower and sessions shorter rules out fatigue as a probable factor in these results.

To summarize, current evidence suggests that stimulus equivalence is less likely to emerge under speed conditions. The two previously mentioned studies probably represent the closest analogues to a fluency training condition in an equivalence paradigm to date, as speed and accuracy were combined. This suggests that this type of emergent performance is severely limited by the introduction of speed into accurate responding. Current evidence from fluency training suggests that speed can increase the probability of retention of learned skills, while the evidence for other fluency indicators is limited. It is possible some fluency indicators are differentially influenced under speed conditions. Some trade-off between retention and application (or equivalence) might even occur under conditions of speed. Further complimentary data will hopefully result from basic and applied behavioral research.

References

Andronis, P. T., Layng, T. V. J., & Goldiamond, I. (1997). Contingency adduction of "symbolic aggression" by pigeons. *Analysis of Verbal Behavior, 14*, 5-17. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2748571/

Arntzen, E. (2011). Comments on Sidman's remarks. *European Journal of Behavior Analysis, 12*, 351-353.

Arntzen, E. (2012). Training and testing parameters in formation of stimulus equivalence: Methodological issues. *European Journal of Behavior Analysis, 13*, 123-135.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record, 60*, 437-462.

Arntzen, E., & Haugland, S. (2012). Titration of limited hold to comparison in conditional discrimination training and stimulus equivalence testing. *The Psychological Record, 62*.

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures versus abstract stimuli as nodes. *The Psychological Record, 60*, 659-678. Retrieved from http://opensiuc.lib.siu.edu/tpr/vol60/iss4/8/

Baron, A. (1985). Measurement scales and the age-complexity hypothesis. *Experimental Aging Research, 11*(4), 193-199. doi: 10.1080/03610738508259187

Bentall, R. P., Jones, R. M., & Dickins, D. W. (1998). Errors and response latencies as a function of nodal distance in 5-member equivalence classes. *The Psychological Record, 49*, 93-115. Retrieved from http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1452&context=tpr

Berens, K. N., Boyce, T. E., Berens, N. M., & Kenzer, A. L. (2003). A technology for evaluating relations between response frequency and academic performance outcome. *Journal of Precision Teaching & Celeration, 19*, 20-34.

Binder, C. (1993). Behavioral fluency: A new paradigm. *Educational Technology*, 8-14. Retrieved from http://www.fluency.org/#Publications%20anchor

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst 19*(2), 163-197. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733609/

Binder, C. (2003). Doesn't everybody need fluency? *Performance Improvement*, 20-28. Retrieved from http://www.fluency.org/Need_Fluency.pdf

Binder, C. (2004). A refocus on response-rate measurement: Comment on Doughty, Chase, and O'Shields (2004). *The Behavior Analyst, 27*, 281-286. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755394/

Binder, C. (2010). Building fluent performance: Measuring response rate and multiplying response opportunities. *The Behavior Analyst Today, 11*, 214-225. Retrieved from http://www.eric.ed.gov/PDFS/EJ937632.pdf

Binder, C., Haughton, E., & Bateman, B. (2002). Fluency: Achieving true mastery in the learning process, from http://www.fluency.org/Binder_Haughton_Bateman.pdf

Binder, C., & Watkins, C. L. (1990). Precision teaching and direct instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3*, 74-96. Retrieved from http://www.fluency.org/PT_DI.pdf

Brady, K. K., & Kubina, R. M. (2010). Endurance of multiplication fact fluency for students with attention deficit hyperactivity disorder. *Behavior Modification, 34*. doi: 10.1177/0145445510361331

Brooks, P. (1995). A comparison of the effectiveness of different teaching strategies in teaching spelling to a student with severe specific difficulties/dyslexia. *Educational and Child Psychology, 12*, 80-88.

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13*, 140-163.

Carriere, J. S. A., Cheyne, J. A., Solman, G. J. F., & Smilek, D. (2010). Age trends for failures of sustained attention. *Psychology and Aging, 25*(3), 569-274. doi: 10.1037/a0019363.

Catania, A. C. (2007). *Learning* (4th ed.). New York: Sloan Publishing.

Chapman, S. S., Ewing, C. B., & Mozzoni, M. P. (2005). Precision teaching and fluency training across cognitive, physical, and academic tasks in children with traumatic brain injury: A multiple baseline study *Behavioral Interventions, 20*, 37-49. doi: 10.1002/bin.168.

Chase, P. N., Doughty, S. S., & O'Shields, E. M. (2005). Focus on response rate is important but not sufficient: A reply. *The Behavior Analyst, 28*, 163-168. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755374/

Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making math easier for pupils and teachers. *Educational Psychology in Practice, 16*. doi: 10.1080/02667360020006372

Dougherty, K. M., & Johnston, J. M. (1996). Overlearning, fluency, and automaticity. *The Behavior Analyst, 19*, 289-292. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733607/

Doughty, S. S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *The Behavior Analyst, 27*(1), 7-23.

Downer, A. C. (2007). The national literacy strategy sight recognition programme implemented by teaching assistants: A precision teaching approach. *Educational Psychology in Practice, 23*, 129-143. doi: 10.1080/02667360701320820

Evans, S. S., & Evans, W. H. (1985). Frequencies that ensure skill competency. *Journal of Precision Teaching, 2*, 25-30.

Evans, S. S., Merger, C. D., & Ben, W. H. (1983). The relationship of frequency to subsequent skill acquisition. *Journal of Precision Teaching, 4*, 28-34.

Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64*, 129-145. doi: 10.1901/jeab.1995.64-129.

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317-332. doi: 10.1901/jeab.1987.48-317.

Fox, E. J., & Ghezzi, P. M. (2003). Effects of computer-based fluency training on concept formation. *Journal of Behavioral Education, 12*, 1-21. doi: 10.1023/A:1022373304577.

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

Greer, R. D., & McDonough, S. H. (1999). Is the learn unit a fundamental measure of pedagogy? *The Behavior Analyst, 22*, 5-16. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2731322/

Heinicke, M. R., Carr, J. E., & LeBlanc, L. A. (2010). On the use of fluency training in the behavioral treatment of autism: A commentary. *The Behavior Analyst, 33*, 223-229. Retrieved from http://www.auburn.edu/~jec0018/pub/HCLS2010.pdf

Holding, E., Bray, M. A., & Kehle, T. J. (2011). Does speed matter? A comparison of the effectiveness of fluency and discrete trial training for teaching noun labels to children with autism. *Psychology in the Schools, 48*, 166-183. doi: 10.1002/pits.20535.

Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record, 50*, 305-337.

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

Imam, A. A. (2003). Assessing transfer of response speed and nodality via conditional discriminations. *Experimental Analysis of Human Behavior Bulletin, 21*, 1-7.

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107-124. doi: 10.1901/jeab.2006.58-04.

Ivarie, J. J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *Remedial and Special Education, 7*, 25-30. doi: 10.1177/074193258600700506.

Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst, 19*, 281-288. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733608/pdf/behavan00020-0139.pdf

Kubina, R. M. (2005). The relations among fluency, rate building, and practice: A response to Doughty, Chase, and O'Shields (2004). *The Behavior Analyst, 28*, 73-76. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755345/

Kubina, R. M., Ward, M. C., & Mozzoni, M. P. (2000). Helping one person at a time: Precision teaching and traumatic brain injury rehabilitation. *Behavioral Interventions, 15*, 189-203. doi: 10.1002/1099-078X.

Kubina, R. M., & Wolfe, P. (2005). Potential applications of behavioral fluency for students with autism. *Exceptionality, 13*, 35-44. Retrieved from http://www.precisionteachingresource.net/08.pdf

Kubina, R. M., & Yurich, K. K. L. (2008). Developing behavioral fluency for students with autism: A guide for parents and teachers. *Intervention in School and Clinic, 44*, 131-138. doi: 10.1177/1053451208326054.

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2004). Engineering discovery learning: The contingency adduction of some precursors of textual responding in a beginning program. *Analysis of Verbal Behavior, 20*, 99-109.

Lindsley, O. R. (1996a). The four free-operant freedoms. *The Behavior Analyst, 19*, 199-210. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733618/

Lindsley, O. R. (1996b). Is fluency free-operant response-response chaining? *The Behavior Analyst, 19*, 211-224.

McDowell, C., & Keenan, M. (2001). Developing fluency and endurance in a child diagnosed with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis, 34*, 345-348. doi: 10.1901/jaba.2001.34-345.

Merbitz, C., Vieitez, D., Merbitz, N. H., & Binder, C. (2004). Precision teaching: Applications in education and beyond. In D. J. Moran & R. W. Malott (Eds.), *Evidence-based educational methods* (pp. 63-80). Amsterdam: Elsevier Academic Press.

Myerson, J., Robertson, S., & Hale, S. (2007). Aging and intraindividual variability in performance: Analysis of response time distributions. *Journal of the Experimental Analysis of Behavior, 88*, 319-337. doi: 10.1901/jeab.2007.88-319.

Olander, C. P., Collins, D. L., McArthur, D. L., Watts, R. O., & McDade, C. E. (1986). Retention among college students: A comparison of traditional versus precision teaching. *Journal of Precision Teaching, 6*, 80-82.

Péladeau, N., Forget, J., & Gagné, F. (2003). Effect of paced and unpaced practice on skill application and retention: How much is enough? *American Educational Research Journal, 40*, 769-801.

Potts, L., Eshleman, J. W., & Cooper, J. O. (1993). Ogden R. Lindsley and the historical development of precision teaching. *The Behavior Analyst, 16*, 177-189.

Shirley, M. J., & Pennypacker, H. S. (1994). The effects of performance criteria on learning and retention of spelling words. *Journal of Precision Teaching, 1*, 73-86.

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15-27). Reno, NV: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research history*. Boston: Authors Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127-146. doi: 10.1901/jeab.2000.74-127.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5-22. doi: 10.1901/jeab.1982.37-5.

Singer-Dudek, J., & Greer, R. D. (2005). A long-term analysis of the relationship between fluency and the training and maintenance of complex math skills. *The Psychological Record, 55*, 361-376.

Spangler, R. S., & Hankins, N. E. (1975). Comparison of two evaluative procedures on retention by college students. *Psychological Reports, 36*, 613-614. doi: 10.2466/pr0.1975.36.2.613.

Spencer, T. J., & Chase, P. N. (1996). Speed analysis and stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643-659. doi: 10.1901/jeab.1996.65-643.

Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior, 85*, 349-369. doi: 10.1901/jeab.2006.107-04.

Twarek, M., Cihon, T., & Eshleman, J. (2010). The effects of fluent levels of big 6 + 6 skill elements on functional motor skills with children with autism. *Behavioral Interventions, 25*, 275-293. doi: 10.1002/bin.317.

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125-144. doi: 10.1901/jeab.1988.50-125.

Equivalence Class Formation in Accuracy or Speed Conditions:

Immediate Emergence, Adduction and Retention

Pétur Ingi Pétursson

Learning in Complex Systems, MALKL 5000, Master Thesis

Oslo- and Akershus University College

Abstract

Three pairs of participants were matched from a pool of 11 based on performance in an MTS pretraining. One participant from each pair completed either a speed or an accuracy condition, in an MTS training using an LS structure for establishing potentially three 3-member classes. Initially, the speed participants completed training with a lowering LH titration on comparison stimuli. The accuracy participants were yoked to a trial number of the speed participants with whom they were matched. In Phase 2, the speed participants completed MTS training with a fixed 1000 ms LH, and the accuracy participants were not yoked. In Phase 3 the probabilities of feedback were gradually reduced to 0 %. Phase 4 introduced tests for directly trained and emergent relations and adduction. In Phase 5, a two-week retention test was administered. The results indicated that one out of three speed participants, compared to all the accuracy participants, responded according to equivalence in the first testing, but adduction was nearly equal for all participants. On a two-week retention test, no accuracy participant and two speed participants responded according to equivalence, while adduction was equal between participants. All speed participants had a higher number of error responses during baseline and completed a higher number of acquisition trials. Inverse reaction time (InvRT) to comparisons did not vary as a function of trial type.

Key words: stimulus equivalence, speed, limited hold, titration, immediate emergence, retention, adduction, reaction time.

Equivalence Class Formation in Accuracy or Speed Conditions:
Immediate Emergence, Adduction and Retention

A conditional discrimination training of arbitrary stimuli using a matching-to-sample (MTS) procedure can make stimuli conditionally related. The stimuli will also be equivalent if correct responding occurs to new, untrained stimulus relations, which show the properties of reflexivity, symmetry and transitivity (Sidman & Tailby, 1982). Reflexivity refers to an identity relation of a stimulus to an identical stimulus (if AB is trained, then AA and BB emerges). Symmetry refers to a relation that shows the interchangeability of two stimuli (if AB is trained, then BA emerges). Transitivity is a relation in which separate stimulus pairs share a common stimulus, where a relation emerges from the separate pairs (if AB and AC are trained, then BC and CB emerge) (Sidman, 1994). Stimulus equivalence has been a central topic in behavior analysis, spanning more than three decades of research (Green & Saunders, 1998; Sidman, 1994), and is still going. In a classical study, Sidman and Tailby (1982) trained six participants in AB, AC and DC conditional discriminations between four stimulus sets, A, B, C and D, with three stimuli in each set (A1, A2, A3, B1, B2, etc.). The 12 directly trained conditional discriminations resulted in a total of 18 additional emergent relations showing properties of reflexivity, symmetry and transitivity, a demonstration of the existence of two 3-member classes, ABC and ABD, and one four-member class, ABCD.

Equivalence classes have a certain structural composition. First, stimuli belonging to a class can be categorized as nodes or singles according to how they are related to other stimuli. A node is a stimulus that is related to at least two other stimuli through training. A single is a stimulus that is related to another stimulus or a node through training. For example, when AC emerges from AB and BC training, A is a single and B is a node (Fields & Verhave, 1987). The number of nodes which separate class members has been shown to influence both accuracy and reaction time (RT) or speed of responding, with a higher nodal number resulting in a reduction in both accuracy and a RT or speed to comparison stimuli in tests for emergent relations (Arntzen, Grondahl, & Eilifsen, 2010; Arntzen & Holth, 2000b; Bentall, Jones, & Dickins, 1998; Fields, Landon-Jimenez, Buffington, & Adams, 1995; Wang, Dack, McHugh, & Whelan, 2011), and one study (Spencer & Chase, 1996) reported a reduction in response speed, but not accuracy.

A training structure to refers to the type of repeated systematic arrangement and directionality of stimuli in training. When the sample stimuli always come from a single set (e.g. set A), and comparisons are from other sets (e.g. B, C, D), the training structure is called a one-to-many (OTM), where the sample is always a node. When the sample stimuli come from many different sets and comparison stimuli from a single, set it is called a many-to-one (MTO) structure, where the comparison is always a node (Imam, 2006; Saunders & Green, 1999). With an MTO or OTM structure, stimuli from each can either serve as samples or comparisons in training trials. A different design features stimuli from each set as samples on one type of training trials, but as comparisons on others. For example, in an ABC class, AB relations would be trained until criteria, followed by the BC relations. This has been called a linear series (LS) structure (Saunders & Green, 1999). Compared to MTO and OTM, LS is least probable to result in equivalence class formation, for both three and four member classes (Arntzen, Grondahl, et al., 2010; Arntzen & Hansen, 2011; Arntzen & Holth, 1997, 2000a; Elifsen & Arntzen, 2009). An LS structure facilitates the discovery of experimental variables that influence the probabilities of class establishment, as a ceiling in accuracy often results from the previous two structures (Holth & Arntzen, 2000; Spencer & Chase, 1996).

Stimulus equivalence research has over a period time been concerned with RT, response speed and speed contingencies. This is for different theoretical reasons; this variable has been suggested as an important behavioral indicator to how equivalence classes form (Holth & Arntzen, 2000; Spencer & Chase, 1996), and a measure that might be sensitive to variables involved in establishing equivalence classes and to their structural composition, such as stimulus relatedness and nodal number. In addition, small-scale differences in RT may give an indication of the substitutability of members within a class, which accuracy measures do not capture (Fields et al., 1995).

Spencer and Chase (1996) showed that participants can have near-perfect accuracy in tests for emergent relations, while response latency or speed may continue to vary as a function of other variables. Using an LS-structure, they trained three groups of participants in matching seven sets (A, B, C, D, E and G) of potential three-member classes – a total of 18 relations. The three groups, an instructed group, a queried group and a standard group, were matched on speed based on a non-significant average difference between the groups in training. Speed was defined as Inverse Reaction Time (InvRT) from the presentation of a sample to the selection of a comparison stimulus. While no differences were found on InvRT or accuracy between the groups in testing, a two-way ANOVA showed that on average a difference between response speed variance in participants responding according to equivalence could be accounted for by a difference between baseline and equivalence (58%) and between baseline and transitivity plus equivalence (36.5%). In addition, symmetry was fastest compared to transitivity, while equivalence trials were slowest on average, accounting in total for 31% of the variance of response speed. In all trial types, response speed was lower on average as a function of nodal number, 14% on symmetry, 17% on transitivity and 18% on equivalence respectively, a result believed to indicate an unequal relatedness of class members (Spencer & Chase, 1996). The conversion of RT to InvRT as used in this study stabilizes variance by reducing the distance between greater RT differences and increasing small RT differences (Baron, 1985). This facilitates visual analyses of data, where it is important to detect systematic small-scale differences.

A number of other studies have shown differential RT to trial types. In general, equivalence trials have the slowest RTs (Arntzen, Grondahl, et al., 2010; Arntzen, Halstadtro, Bjerke, & Halstadtro, 2010; Arntzen & Lian, 2010; Bentall et al., 1998; Elifsen & Arntzen, 2009; Holth & Arntzen, 2000; Imam, 2001, 2006; Spencer & Chase, 1996). Some have reported slower RTs in symmetry trials than in baseline trials (e.g. Holth & Arntzen, 2000; Spencer & Chase, 1996). One study showed that RT on equivalence trials reach the level of symmetry trials with continued testing (Arntzen & Lian, 2010), and mixed results have also been reported (e.g. Arntzen & Haugland, 2012; Tomanari, Sidman, Rubio, & Dube, 2006). The effect has occurred under both accuracy and speed conditions. For example Imam (2001, Experiment 2) used a single-subject design for two participants exposed to an equal number of trials in a condition of either accuracy followed by speed or speed followed by accuracy. Three 5-member and three 7-member classes were trained with an LS training structure. Results showed that trial types in testing accounted for 12–15% of the variance in response speed, being fastest on baseline, followed by transitivity and equivalence, with and without a speed contingency in place. One study showed that the difference between relational types ceases to occur when namable, instead of arbitrary, stimuli are used (Bentall, Dickins, & Fox, 1993). Where picture stimuli have been used as nodes the difference between relational types occurs, but to a lesser degree than with abstract stimuli only (Arntzen & Lian, 2010). Abstract stimuli result in longer RTs on tests for transitivity and equivalence (Arntzen, Grondahl, et al., 2010; Bentall et al., 1993), or only on equivalence trials (Arntzen & Lian, 2010). While it has repeatedly been shown that nameable stimuli facilitate stimulus equivalence formation (Arntzen, 2004; Arntzen & Lian, 2010; Bentall et al., 1993; Holth & Arntzen, 1998a), one study has shown that meaningful nonsense-syllable stimuli hinder equivalence more than non-meaningful syllables (Lyddy, Barnes-Holmes, & Hampson, 2000).

Another general finding is that RTs become slower from the last baseline trials to the first test trials (Arntzen, 2004; Arntzen, Grondahl, et al., 2010; Arntzen, Halstadtro, et al., 2010; Arntzen & Holth, 1997, 2000a; Arntzen & Lian, 2010; Holth & Arntzen, 1998a, 1998b; Hove, 2003; Imam, 2001; Spencer & Chase, 1996; Wulfert & Hayes, 1988) and become faster with repeated testing (Arntzen, 2004; Holth & Arntzen, 1998a, 2000). This difference is more pronounced with all abstract stimuli, than when one meaningful stimulus is used as a class member (Arntzen & Lian, 2010; Holth & Arntzen, 1998a). An interpretation of the difference in RT between baseline and testing has posited a mediating naming response (Horne & Lowe, 1996) or a precurrent problem-solving behavior, particularly following an LS training protocol, where a participant only has experience of each sample with one comparison, and is exposed to a novel stimulus arrangement in testing (Holth & Arntzen, 2000). A general hypothesis would be if slower or faster RTs in baseline can predict stimulus equivalence class formation. Some studies have reported limited predictive validity of RTs to comparison stimuli in baseline for the establishment of equivalence classes (e.g. Arntzen & Holth, 1997). However one study demonstrated a relationship between RTs and symmetric responding and certain response patterns. Holth and Arntzen (2000) trained 10 participants in matching three 3-member classes with a LS training structure followed by testing with a simultaneous protocol, and found that of the four participants who established equivalence, the three that had the total fastest RT to comparison stimuli were more likely have responded correctly on symmetry trials and have consistent response patterns on tests for equivalence. In general, consistent rather than nonsystematic responding had a lower RT on average.

A speed contingency can introduce a possible barrier to any precurrent behavior, as faster responding can limit the time available to engage in this type of behavior. For example, in Holth and Arntzen (2000, Experiment 3), 10 participants were trained using an OTM structure establishing three 3-member classes, with a 2-s LH to comparison stimuli during training and testing, where only correct responses within the time limit received positive feedback. Under these conditions, only five reached MTS training criteria, none of them responding according to equivalence under a simultaneous testing protocol. Whether precurrent behavior or naming occurs or is hindered, the manipulation of environmental variables such as an LH is important for its own sake: It adds to the understanding of what conditions are more or less likely to form equivalence classes (Tomanari et al., 2006). A comparison of speed and accuracy conditions can, for example, be important for future instructional technology with a foundation in basic research.

Imam (2001) provided data from two participants showing that a speed contingency results in a decline in response accuracy on transitivity and equivalence trial types, compared to directly trained relations. However, in an accuracy contingency they maintained high accuracy on the same relations. A similar trade-off between speed and accuracy was also reported in systematic replications with four participants, training potentially seven-member classes and three different training and testing sequences. In that case, however, the majority of errors occurred when participants were unable to respond within a given time limit (Imam, 2006). While many researchers have introduced a speed contingency with a fixed LH, a gradual introduction of an LH to sample or comparison stimuli in an MTS task has also been attempted recently. Tomanari et al. (2006) gradually lowered an LH to sample and comparison stimuli in an MTS-task for establishing potential four-member classes, using an OTM training structure. The five participants progressed through lowering LHs until a criterion accuracy of 95% occurred within an LH of 0.4–0.5 s on samples and 1.2-s to comparisons. Three of the five participants then responded in accordance with equivalence. The participants' RTs were equal in training and testing trials, but for all participants response speed tended to be marginally slower on average (a 0.2–0.3-s difference) when responding to comparisons in symmetry and equivalence trials, compared to baseline trials.

Marginal, but consistently higher RTs were also reported for equivalence vs. symmetry trials. Arntzen and Haugland (2012) conducted a systematic replication of this study, with the modification that a computer titrated the LH to comparison stimuli, and they used three- instead of four-member classes and shorter MTS training sessions. Only one of these five participants responded in accordance to equivalence under a 2.5-s LH, and no RT differences to any trial type could be documented.

Sidman (1992) suggested that the immediacy of equivalence may be a measure sensitive to whether experimental or uncontrollable variables are responsible for positive test outcomes. However the degree of immediate emergence can be a function of experimenter-controlled variables, such as training conditions and testing protocols (e.g. Fields & Garruto, 2009) or previously consistent MTS responding (Holth & Arntzen, 1998a). Another suggestion may be the stability of equivalence classes. For example in a study by Saunders, Wacther and Spradlin (1988), three out of four participants achieved above 90% accuracy on tests for directly trained and emergent relations after up to 206 days without training, and Rehfeldt and Hayes (2000) obtained similar results after three months without training. Symmetric relations are more likely to be retained after two to three months than equivalence relations, but symmetry is retained for all participants who have equivalence intact (as cited in Dymond & Rehfeldt, 2001, p. 10).

In general, the lower probability of equivalence classes being established with a speed contingency (Arntzen & Haugland, 2012; Holth & Arntzen, 2000; Imam, 2001, 2006; Tomanari et al., 2006), particularly with an LS structure (Holth & Arntzen, 2000; Imam, 2001, 2006), suggests that a speed condition limits equivalence formation. In addition, the high trial number needed to reach criteria (Arntzen & Haugland, 2012; Imam, 2001; Tomanari et al., 2006) suggests that speed is a more difficult MTS training condition than an accuracy contingency. The main purpose of this study was to investigate further whether speed or accuracy conditions result in a differential likelihood of responding according to equivalence, by matching speed and accuracy participants. This study also included few test trials, to be able to assess whether speed or accuracy influences the immediate emergence of equivalence, which could simultaneously reduce the probabilities of effects carrying over from one testing condition to another. A second research question was whether different probabilities of adduction from the emergent relations would occur from speed or accuracy. This would require a test for the demonstration of a recombination of the emergent relations into novel performances (e.g. (Andronis, Layng, & Goldiamond, 1997). A third question was if there would be a difference in the stability of the emergent relations and adduction, by administering a retention test following two weeks without practice. A fourth purpose was to yoke speed and accuracy participants during one phase of baseline, making it possible to do a closer inspection how either condition influences acquisition of directly trained relations. An additional research question was whether the typically reported response speed or RT differences between trial types, or training and test trials would appear following either speed or accuracy conditions.

Method

Participants

Through a college campus, 11 participants were recruited. They ranged in age from 22 to 30. The six participants who completed all the experimental phases ranged from 22–29 (M_{age} = 25). Three were male and three were female; two were Masters students, one held a Masters degree, two were Bachelor students, and one a Doctoral student.

Setting

The experiment took place in Oslo and Akershus University College in the Laboratory of Complex Human Behavior Studies. Participants were seated in booths about 1.75 × 5 m in diameter, in front of a 45 × 90 cm table in quiet environment. The participants faced a wall or a window with drawn curtains. The booths were situated in two different housing locations affiliated with the laboratory; approximately 25 m^2 in size and the other about 20 m^2.

Instruments

A custom made-program, Matching to Sample© (second edition), was used to run the experiment. The program was run on a Hewlett Packard HP Compaq Nc 6320 PC on a 32-bit Windows 7 Professional. The screen of the laptop was 15”, 16:9 and with 1400X1050 resolution. A Magic Touch KTMT-1500-USB touch screen was mounted onto the screen. A Magic Touch v.2.21 for Windows 7 driver was used to run the touch screen.

Stimuli

For the Pretraining phase, three numbers and three nonsense syllables were used as stimuli, approximately 2 × 3 cm in size. Six arbitrary 2.5 × 2.5 cm stimuli and three numbers were used in the remaining phases (see Figures 1 and 2).

Experimental Design

A matched design with a partial yoking of accuracy to speed was used for this experiment. (E. Arntzen, personal communication, February 2, 2011). The first two participants were assigned to a speed contingency. The subsequent participants were then matched to one speed participant and assigned to an accuracy contingency, only if one was within ±12 trials to criterion in the Pretraining Phase. If the next participant could not be matched to a speed participant, this participant also completed a speed condition. Participants who followed were successively matched based on the trial number needed to complete the Pretraining Phase. One participant was unable to complete the Pretraining, and did not move further to the experimental phases. One speed participant was unable to complete Phase 1, and another speed participant did not complete Phase 2. Three participants, who could be matched to one of the speed participants, then entered the experiment at various stages, and were assigned to an accuracy condition. Two of the remaining speed participants dropped out of the analysis as they could not be matched. This resulted in three pairs of matched participants.

General Procedure

All participants read an information sheet, which explained the broad goals of the research conducted at the laboratory, although the purpose of the experiment was not mentioned and stimulus equivalence was neither defined nor explained. In addition, their rights as research participants were explained, and they were informed that they could withdraw their participation should they wish. An informed consent was then read and signed. Each session lasted from about 1.5 to 3.5 hrs. The sessions started with instructions that were separate for each condition of the experiment, as described below. When the participants finished reading the instructions, they pressed a square saying “begin” on the bottom of the touch screen to begin the first trial. A trial started with presentation of a single sample stimulus in the center of the screen. Touching the sample stimulus made it disappear and three comparison stimuli appear simultaneously, with a 0-s delay. The program determined the positioning of the comparison stimuli randomly from trial to trial. The comparison stimuli appeared in a circular layout, 10 mm from the sample stimulus. Choosing one of the sample stimuli gave a 500 ms feedback; correct class-consistent responses were followed by “good”, “excellent” etc.; choosing an incorrect comparison stimulus was followed by “wrong”. No consequences resulted from any other type of responses. The bottom of the screen presented a count of correct responses. Reaction time was recorded based on the interval between touching the sample stimulus and selection of a comparison stimulus, which was transformed into inverse reaction times (InvRT). Intertrial interval (ITI) was set to 1000 ms in all phases, which resulted from the end of the feedback interval, where the screen remained black. No consequences were delivered upon touching the screen during the ITI or presentation of the feedback. Participants completed Phase 1, 2 and 3, which were different for the speed and accuracy participants, while Phases 4 and 5 were identical for both groups. Phases 1–4 were completed in one session, while Phase 5 was completed two weeks later. The experiment succeeded through a simultaneous protocol, in which all relations were first trained (Phases 1–3), and probe trials were presented in a mixed order in separate test blocks (Phases 4 and 5).

Training

Pretraining. This condition was identical for all participants. First, they were presented with the following instructions:

> A stimulus will appear on the middle of the screen. Choose it by pressing on the screen. Two other stimuli will appear. Choose one of these by pressing on the screen in the same manner. If you choose the stimulus we have defined as correct, words like "very good", "excellent" and so on will appear on the screen. If you press the wrong stimulus, the word "wrong" will appear on the screen. At the bottom of the screen the number of correct responses you have made will be counted. Please do your best to get everything right. Good luck! Press start to begin the experiment.

Potentially two 3-member classes of nonsense syllable stimuli (see Figure 1) were then trained with an MTO training structure to perform the following conditional discriminations: A1C1, A2C2, B1C1 and B2C2, with blocks containing 36 trials. Training order was concurrent, so that multiple classes were trained simultaneously, and the trial types were not introduced in a particular order. The training finished after participants had reached a 100% accuracy criterion within a single block. An additional 36-trial training block was added after the accuracy criterion was achieved, after which pretraining ended.

Phase 1. This condition was different for the speed and accuracy participants. The speed participants were presented with the following instructions:

> A stimulus will appear on the middle of the screen. Choose it by pressing on the screen. Three other stimuli will appear. Choose one of these by pressing on the screen in the same manner. If you choose the stimulus we have defined as correct, words like “very good”, “excellent” and so on will appear on the screen. If you press the wrong stimulus, the word “wrong” will appear on the screen. At the bottom of the screen the number of correct responses you have made will be counted. Try to press as quickly and correctly as you can. Good luck! Press start to begin the experiment.

Potentially three 3-member classes were then trained with an LS training structure. The following conditional discriminations were trained: A1B1, A2B2, A3B3, B1C1, B2C2 and B3C3. Trial types were again introduced with a concurrent protocol. Each training block consisted of 18 trials, divided into three six-trial titration blocks. Sample stimuli were presented with a fixed 1000 ms limited hold (LH). A descending titrated LH to comparison stimuli occurred in 100 ms steps. The initial value of the titration was based on the average reaction time of the last five trials in the Pretraining Phase, with 1000 ms added. Progressing through the titration blocks was based on a 100% accuracy criterion in six consecutive trials. Achieving the accuracy criterion reduced the LH one step. If the participants responded incorrectly to the comparison stimuli, or did not respond within the LH, a negative feedback occurred, and a new trial began. Incorrect or correct responses that occurred after the LH ended were recorded separately as missed. The titration continued until the participants reached the accuracy criterion with a 1000 ms LH. This added an extra block of 18 trials of the lowest LH value. Reaching the 94% accuracy criterion on the lowest LH in a single block ended this phase with a five-min break before the next phase started.

The accuracy participants began the same phase with the same instructions as the speed participants, except that the sentence “Try to press as quickly and correctly as you can” was replaced with “Try to press as correctly as you can”. Accuracy participants were then yoked to an equal trial number resulting from one matched speed participant in this phase. These trials appeared in one block, with the LH to sample and comparison stimuli not in place. Other experimental parameters were identical.

Phase 2. The speed participants were presented with the same instructions, stimuli and training protocol as in the previous phase, but the titration was dropped and the sample and comparison stimuli were presented with a fixed 1000 ms LH. Each training block had 18 trials and an accuracy criterion of 94%. Reaching the criterion in one training block triggered the next phase. The accuracy participants were presented with the same instructions as in Phase 1, and were then exposed to a new block of 18 trials without an LH to sample or comparison stimuli and with no yoking in place. Other experimental parameters were identical to those of the speed participants.

Phase 3. This phase introduced blocks of 18 trials, in which the probability of programmed consequences was reduced to 75%, 25% and then 0%, if 94% accuracy was reached in individual blocks. If the 94% accuracy criterion was not reached, this set in a previous fading block. Reaching an accuracy criterion with the 0% probability of programmed consequences (i.e. extinction) immediately started the next phase. The speed participants completed this phase with a 1000 ms LH, and the accuracy participants without an LH.

Testing

Phase 4. This condition was identical for all participants and consisted of one block of 36 mixed-test trials presented in a random order. One exception was Accuracy Participant 6, who completed 54 test trials in the first test. His first 36 test trials were used for analysis. First, the participants were presented with the following instructions:

> A stimulus will appear on the middle of the screen. Choose it by pressing on the screen. Three other stimuli will appear. Choose one of these by pressing on the screen

> in the same manner. Try to get as many correct as possible. Good luck! Press start to begin the experiment.

The 36 test trials included 12 trials for directly trained, 12 trials for symmetry, six trials for transitivity and six trials for equivalence. The trial types involved in tests for symmetry were: B1A1, B2A2, B3A3, C1B1, C2B2, C3B3, for transitivity: A1C1, A2C2, A3C3 and for equivalence: C1A1, C2A2, C3A3. Test types were mixed in a randomly determined order, and each relational type was tested twice. Each test had a 100% accuracy criterion and no feedback occurred at any test trial. Following completion of the last test trial, the participants received a message on the screen that the experiment had ended. Seated in the same booth, the experimenter then administered an adduction test with the verbal instructions: “Please calculate the following addition problems.” The adduction test consisted of an A4 piece of paper showing all possible relational types of symmetry, transitivity and equivalence, totaling 15 relations set up as summation tasks, similar to the one used in Bucklin, Dickinson and Fox (2000) (see Figure A1, Appendix).

Phase 5. This phase was a retention test administered two-three weeks after Phase 4. They were exposed to the same instructions, and completed an identical testing, including the adduction tests, as in Phase 4.

Results

The participants completed the Pretraining in the range from 60 to 144 trials. All participants were matched based on a ±12 trials difference in reaching the 100% accuracy criterion. The first pair, Participants P9056 and P9057, completed Pretraining in 72 and 60 trials; the second pair, Participants P9058 and P9059, in 48 and 60 trials; while in the third pair, Participant P9060 completed in 144 trials and Participant P9061 in 132. The average RT for the speed participants in the last five trials of their Pretraining resulted in initial LH values of 4955, 3927 and 1771 ms respectively, from where the titration started. As shown in Table

1, the speed participants needed 390, 432 and 492 trials to complete the titration in Phase 1. Matched accuracy participants were then yoked to complete the same trial number in this phase. In this phase, the number of correct responses for Pair 1 was 317 vs. 356 for P9056 and P9057, in Pair 2 this was 283 for P9058 vs. 414 for P9059; in Pair 3 this was 319 vs. 461 for P9060 and P9061 respectively. In Phase 1 all accuracy participants had a higher number of correct responses than their matched counterparts, or any speed participant.

Speed and accuracy participants differed substantially in the trial number needed for completion of Phases 2 and 3, where the LH was fixed for the speed participants. In Pair 1, P9056 completed 162 trials, while P9057 completed 72 trials; in Pair 2 P9058 completed 1620 trials, while P9059 completed 72 trials; in Pair 3 P9060 completed 720 trials, while P9061 completed 90. This was a higher trial number than in Phase 1 for P9058 and P9060. A lower trial number was needed for the accuracy participants. Only P9056 had a lower number of missed responses from Phase 1 to Phase 2; the other two speed participants showed an increase in missed responses.

The combined acquisition trial numbers through Phase 1 to 3 were clearly higher for all speed participants than for their matched counterparts. The trial difference in Pair 1 was 462 vs. 428 for P9056 and P9057; in Pair 2 these were 1344 vs. 469 acquisition trials for P9058 and P9059; while in Pair 3, P9060 and P9061 completed 853 vs. 529 acquisition trials respectively. The high number of acquisition trials for Speed Participants P9058 and P9060 can be largely accounted for by trials needed to complete Phase 3, where the probabilities of programmed consequences were reduced.

Table 2 shows the results of all tests for directly trained and emergent relations as well as adduction. During the first testing, P9056 did not respond according to the directly trained or emergent relations. P9057 responded according to symmetry, transitivity and equivalence. The directly trained relations were only intact for P9057. Both participants in Pair 1 scored

100% for accuracy in the test for adduction. On retention, neither participant in Pair 1 responded according to any of the emergent relations, but they continued to have 100% accuracy in the adduction test. In Pair 2, P9058 responded according to symmetry, transitivity and equivalence. P9059 responded according to transitivity and equivalence. Both participants had the directly trained relations intact and had 100% accuracy on adduction. All participants returned retention following exactly 14 days, except P9056 who returned following 28 days, and P9060 after 16 days. During retention, P9058 continued to respond according to symmetry, transitivity and equivalence. P9059 no longer responded according to equivalence or transitivity. For both participants the directly trained relations remained intact, and both scored 100% for accuracy on adduction. In Pair 3, P9060 did not respond according to any of the emergent relations, but the directly trained relations remained intact and the participant showed 100% accuracy in the adduction test. P9061 responded according to symmetry and transitivity, with the directly trained relations intact, and scored 93% for accuracy on the adduction test. On retention, P9060 responded according to symmetry and transitivity with the directly trained relations intact. P9061 continued to respond according to symmetry but no longer responded according to transitivity. Both participants in Pair 3 continued to have the directly trained relation intact and to score 100% in the adduction test.

Figure 3 presents the average InvRT to comparison stimuli in Phase 1, where accuracy participants were yoked to the speed participants. InvRT became gradually higher in Phase 1 for both the speed and accuracy participants. Both the speed and accuracy participants in the first and second pair had on average a lower InvRT than the InvLH values in the last blocks of training, as did P9060, while the InvRT of P9061 was below the InvLH values. The terminal InvRT values (last six trials) were always higher for the speed participants than those of their matched participants, or any accuracy participant. In Pair 1, P9056 had an InvRT of 1.2 compared to 0.8 of P9057, in Pair 2 P9058 had an InvRT of 1.28 compared to

0.93 of P9059, and in Pair 3 P9060 had an InvRT of 1.58 compared to 0.48 InvRT of P9061. However, there is a considerable overlap of these values for Pair 1 (P9056 and P9057), and Pair 2 (P9058 and P9059), throughout the phase, while P9061 was clearly responding faster than P9060 throughout the phase. Overall the accuracy participants showed more variability of InvRT than the speed participants. Variability of InvRT was stable for Speed Participants P9056 and P9058, and lower compared with that of their matched Accuracy Participants P9057 and P9059, while P9061 showed similar variability to P9060.

Figure 4 shows the InvRT throughout the test trials during the first testing and retention. P9056 showed more stable responding than P9057 in the first test, and was faster in 24 out of 36 trials. In the retention test the values overlap considerably, but P9056 has more stable responding in the first 21 trials and was faster on 20/36 trials. P9058 and P9059 had a similar variability on the first testing, with considerable overlap, but P9059 was faster than P9058 in 17/36 test trials. However, on the retention test P9058 was faster on 27/36 test trials. P9061 had faster responding on 30/36 test trials during the first test and 28/36 trials during the retention test. P9060 continued to show slightly more variability during the first testing, while the variability was similar for both participants during retention. On average, most participants showed only a marginal difference in InvRT between the first and second testing. This was a 0.5% increase for P9056, a 2.9% increase for P9057, equal for P9058, a 2.3% increase for P9059, a 0.9% increase for P9060, and a 3% decrease for P9061.

The difference in InvRT between the last five training trials and the first five test trials (see Figure 5) was 0.46 vs. 0.93 for P9056 and P9057, 0.55 vs. 0.17 for P9058 and P9059, and a 0.39 vs. 0.32 for P9060 and P9061 respectively. The first five retention test trials were slower on average compared to the first testing. The exception was P9058, who was slightly faster, while P9057 and P9061 had a marginal slowdown.

During the first testing, average InvRT in separate trial types was always higher for the speed participants than for their matched counterparts (see Figure 6), although the differences between P9058 and P9059 were marginal in the directly trained and symmetry trials. In addition, the speed participants were always faster than any accuracy participant. Two out of three speed participants had roughly double the speed of their matched counterparts in symmetry and equivalence (Pairs 1 and 2 respectively). A comparison of trial types within participants shows that the directly trained relations were fastest for all accuracy participants and two out of three speed participants (P9056 and P9058). P9060 accomplished the transitivity trials fastest, but for all other participants these were the third fastest. No other consistent differences between trial types were found during the first testing.

On the retention test, the speed participants continued to respond faster than their matched counterparts in all trial types, except that P9057 was faster than P9056 in the transitivity trials, and nearly equal in the directly trained trials. Comparing the matched pairs, only the average InvRT in the equivalence trials continued to be higher for the speed participants. All participants responded slower in the directly trained relations on retention compared to the first testing. P9056 responded slower in all trial types during retention. All accuracy participants slowed down their responding in equivalence trials during retention, while two out of three speed participants had an increased InvRT on the same trials. No other consistent InvRT differences were found between testing and retention.

Discussion

The main purpose of this study was whether a speed or an accuracy condition results in a differential likelihood of responding according to equivalence. The speed condition resulted in one out of three speed participants responding according to equivalence, compared to three accuracy participants. The low trial number in testing indicates that a speed condition reduces the immediate emergence of equivalence.

A second purpose was to test whether different probabilities of adduction occur following MTS training in either speed or accuracy. All participants had 93–100% accuracy in math problems containing all trial types of symmetry, equivalence and transitivity, suggesting that the likelihood of adduction is equal following either accuracy or speed.

A third purpose was to investigate retention. Two speed participants responded according to equivalence during the two-week retention, compared to no accuracy participants. The only participant to show perfect stability of all the emergent relations over time was Speed Participant P9058, who responded according to symmetry, transitivity, and equivalence during the first test and on a two-week retention test. The only accuracy participant to initially respond according to all the respective relations did not show stability of the emergent relations over time. Symmetric relations only remained stable for accuracy participant P9061, compared to two speed participants P9058 and P9060. These results have to be considered somewhat mixed. However they agree with the results of Rehfeldt and Hayes (2000) in that more subjects show retention of symmetry than equivalence or transitivity, and that retention of equivalence co-occurs with retention of symmetry. It is also noteworthy that the speed participant who did not respond according to the directly trained- and emergent relations in both testing phases also had the lowest number of acquisition trials of all speed participants, so it is possible that stable equivalence classes requires overtraining of the directly trained relations. In this study emergence of equivalence and symmetry occurred for Participant 9060 from the first testing to retention. This indicates that delayed emergence may occur even after a considerable period without exposure to experimental stimuli. It is unclear why this occurred but his responding in the initial symmetry trials were consistent, while the equivalence test trials were not consistent (50% accuracy). Holth and Arntzen (1998a) found that with a group of 50 participants, there was a tendency for delayed emergence of consistent responding with repeated testing. Earlier research has also shown that accuracy on directly trained and emergent relations can improve with repeated testing during retention for some participants (e.g. Rehfeldt & Root, 2004).

Retention of adduction continued during the retention test, where all participants had 100% accuracy. This serves as a replication of results during the first testing; speed or accuracy do not differentially influence the probabilities of adduction. As multiple stimulus relations were presented at once during the adduction test however, it is difficult to say whether adduction is an additional demonstration of responding according to the emergent relations. However, all participants reached their solutions with the same strategy: They wrote down the numbers related to the arbitrary stimuli before giving their answer. As the intermediate responding produced stimuli which occasioned the solution, this would be an example of problem solving, a behavior which may have facilitated adduction from of the emergent relations (e.g. Holth & Arntzen, 1998b; Holth & Arntzen, 2000). Similar results were reported by Rehfeldt and Hayes (2000), where equivalence did not emerge in initial MTS testing, but did emerge in later generalization tests. The current results also suggest that class establishment through MTS training is not limited to tests bound by specific experimental operanda; on the contrary it may produce novel performance outside of the experimental situation. The adduction test can however pose a threat to experimental control, as some practice may have been involved in solving the problems, which might explain why some participants achieved equal or higher accuracy in later MTS retention tests. However, the fact that neither stability nor delayed emergence of the emergent relations occurred for any of the accuracy participants after the retention interval makes this an unlikely explanation.

A fourth purpose was to compare a difference between speed and accuracy by initially yoking accuracy to speed during baseline. The proportionally higher number of error responses occurring for all speed participants during Phase 1 reveals that a speed condition increases the variation of response accuracy in baseline, more so than an accuracy condition. A trade-off between speed and accuracy (Imam, 2001) on tests for emergent relations has been reported (e.g. Holth & Arntzen, 2000; Imam, 2001, 2003). Based on the yoking in Phase 1 of this experiment, this trade-off also occurs during acquisition of the directly trained relations and remains even given no fast-response requirements in testing. These results suggest that the source of the trade-off is in the reinforcement contingency (see Imam, 2006). When the yoking condition was removed, the speed condition resulted in two of the three speed participants using a substantially higher number of trials to reach criterion performance in Phase 2. Similarly, in the Imam (2001, experiment 1) study, participants who completed a speed contingency were reported to need a higher number of blocks and additional maintenance blocks to reach criteria. In the study by Tomanari et al., participants needed up to over 20,000 trials to reach criteria, and the study by Arntzen and Haugland up to 2934 trials, while in this study the highest number of acquisition trials was 1344. This difference may be accounted for by fact that all arbitrary stimuli were used in the two previous studies, and it is clear that four instead of three classes as used in the study of Tomanari et al. resulted in a higher number of acquisition trials. In addition, the high number of trials that resulted from Participants P9058 and P9060 in this study can largely be accounted for by the large number of fading trials. It is possible that a successful consequence fading requires a behavioral event, such as naming to facilitate a correct comparison selection (e.g. Horne & Lowe, 1996). One interpretation is that the speed participants had difficulty reaching accuracy criteria during the fading of consequences as the fast responding limited the time gap available for this behavior to occur.

A fifth question was whether an InvRT would differ on trial types and training and test trials. Earlier studies have shown a variation in RT as a function of relational types. For example, Spencer and Chase found that response speed was an inverse function of relational type, with baseline being fastest, followed by symmetry and equivalence, both during speed and accuracy conditions. Imam (2001) showed this occurring both in a speed and an accuracy contingency. No such patterns in InvRT occurred for either speed or accuracy in this study. In fact, the transitivity trials were nearly as fast as the baseline trials for most participants. In addition, equivalence trials were nearly equal in InvRT to the symmetry trials. These results are unexpected as the LS structure results in the transitivity and equivalence tests having a one-node distance (e.g. Fields, Landon-Jimenez, Buffington, & Adams, 1995). This pattern continued in retention, with minimal slowdown in InvRT occurring for some participants. The difference from earlier studies may be accounted for by the higher number of testing trials they used (e.g. Imam, 2001, 2003; Spencer & Chase, 1996). In addition, this study used a nameable stimulus, which has been shown to eliminate the differences between RT on relational types (Bentall, Dickins, & Fox, 1993). However, both Tomanari et al. and Arntzen and Haugland used entirely arbitrary stimuli and applied a higher number of testing trials than this study, but neither reported differences between relational types. It therefore seems unlikely that the nameable stimulus used in this study reduced the trial type difference. However, some studies (e.g. Imam, 2001; Spencer & Chase, 1996), which have reported a different substitutability of class members based on speed differences between relational types, have analyzed RT or speed of highly accurate responding. Looking at the speed of the five participants who showed high accuracy in tests for emergent relations in this study, neither the study by Arntzen and Haugland nor that of Tomanari et al. reveals any systematic pattern, which suggests that class members were not differentially related as a function of trial type. It is possible that the stepwise reduction in LH, or overtraining balances the effects of InvRT over time.

Current results showed that the difference between the last training trials and the first test trials was more pronounced for the speed than the accuracy participants. This difference can be traced to a differential reinforcement of fast responding during baseline. One behavioral interpretation of this difference is that the initially slower responding during the first test trials indicate precurrent problem solving (e.g. Holth & Arntzen, 2000). The fast response requirements during baseline may have hindered emergence of equivalence during the initial test for the speed participants. However, this contrasts with the results reported by Arntzen and Haugland and Tomanari et al, where the difference was minimal. One additional finding was that the speed condition resulted in faster responding in the acquisition trials of the speed participants in Phases 1 and 2 of this experiment, and in the majority of test trials the speed participants continued to respond faster, which continued following a two-week retention interval. As these MTS performances were clearly contingency shaped, this dimension of responding according to the emergent relations will have to be traced to the reinforcement contingency (see Sidman, 2000). Imam (2003) obtained similar results with a participant who continued to show fast responding on a transfer test to members of an equivalence class; a class not previously trained with a speed contingency.

Variability of InvRT was higher for the accuracy participants in Phase 1. Participants P9056 and P9058 had similar variability in InvRT, but they started out at with a high LH. P9060, who started out with a substantially lower LH than the other speed participants, had the highest variability in InvRT in this phase; the difference between him and the participant to whom he was yoked is most obvious. His matched counterpart also showed a similar variability, although the InvRT values are much lower. It is possible that different variation of speed or RT indicates the stability in the formation of stimulus classes.

Some similarities and differences from similar studies have to be considered. Unlike many previously mentioned studies, which have studied speed conditions, this study introduced tests without an LH being in place. In one part of their experiment, Holth and Arntzen (2000) included a testing condition similar to the one used in the current study. In

Experiment 3, they exposed participants to an MTO training with a 2-s LH where one testing condition was presented without the LH in place. Here three out of five participants showed response patterns consistent with equivalence. When the participants were then tested with an LH in place, consistent equivalence response patterns occurred for only one of those participants. Hence, equivalence formation seems to become more severely limed if time limits to respond are present in testing. For example, Imam (2001, 2003), Tomanari et al. and Arntzen and Haugland presented an LH in testing which limited equivalence for most participants. In Tomanari et al. study, testing was conducted under the terminal LH values that resulted from training, which was between 0.4 and 0.5 s for samples and 1.2 and 1.3 s for comparisons. Arntzen and Haugland used a 2.5-s LH on comparisons in testing, and Holth and Arntzen used a 2-s LH in testing. A previous interpretation by Holth and Arntzen (1998b, 2000), that novel stimulus arrangements in testing require precurrent problem solving to which speed may be a barrier, is given further support by the finding that even previously shaped fast responding may limit such an intermediate behavior.

This study was conducted in two sessions, unlike previously mentioned studies on speed, which ran participants through multiple sessions. Another difference is that the third member of each class was a nameable stimulus, which may have facilitated the completion of MTS training in just one session (e.g. Arntzen & Lian, 2010). However, finishing the experimental phases in two sessions provided some safeguard against external variables influencing the results and a quicker comparison of speed and accuracy in an MTS paradigm. This study also included only six test trials for equivalence and transitivity, along with 12 symmetry and baseline trials, which meant that the accuracy criteria for responding according to the emergent relations were different from those in earlier studies. Keeping the test trial number low made it possible to evaluate whether immediate emergence would occur, but also reduced possible practice with the emergent relations carrying over to the adduction tests. In addition, this study used a concurrent training order, while the Arntzen and Haugland study and Tomanari et al's. study used a serialized training order, in which individual conditional relations were trained to criteria, until mixed with other relations. The study by Haugland and Arntzen titrated down to 1200 ms, and used an 80% accuracy criterion for finishing both individual titration blocks and across the last blocks of training. Tomanari et al. used a 95% accuracy criterion across two consecutive blocks. The relatively stricter accuracy criterion of 94% in this study and Tomanari may have contributed to the higher number of participants finally responding according to equivalence. Another difference is that the ITI in this study was 1 s, whereas Tomanari et al. used a 0.4-s ITI and Arntzen and Haugland a 0.5-s one, so the overall pace may have been faster in the previous studies, while in this study faster responding occurred during the final trials of the titration. The previous study of Arntzen and Haugland to employ titration used an OTM structure in training the conditional discriminations, while this study used an MTO pre-training, followed by an LS structure for the experimental stimuli. In addition, with the exception of Arntzen and Haugland's study, previous studies on equivalence and speed contingencies have not used the fading of consequences before starting an extinction phase. For example, Tomanari et al. used a training phase without differential consequences, which was later mixed with probes for emergent relations. Like Arntzen and Haugland, this study presented a mixed test of all emergent relations following baseline training. It also used a considerably higher start value of the LH than previous studies. The fact that more speed participants finally responded according to equivalence in this than in previous studies may partly be accounted for by this difference. Whether the difference between speed and accuracy is due to the total higher number of trials of speed participants during training or the actual speed condition can only be answered by further experimentation. Yoking accuracy to speed through all experimental phases is one possibility. It would be interesting to determine how this influences adduction.

It is possible that the relatively stable InvRT during the titration in this study indicates the stability of class formation. It would be interesting to see whether the introduction a titrated LH or a fixed LH results in some difference in the stability of equivalence classes. Finally, further research may determine whether speed and accuracy differentially influence the immediate or delayed emergence of stimulus equivalence.

References

Andronis, P. T., Layng, T. V. J., & Goldiamond, I. (1997). Contingency adduction of "symbolic aggression" by pigeons. *Analysis of Verbal Behavior, 14*, 5-17. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2748571/

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*, 275-291.

Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *The Psychological Record, 60*, 437-462.

Arntzen, E., Halstadtro, L.-B., Bjerke, E., & Halstadtro, M. (2010). Training and testing music skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions, 25*(2), 129-143. doi: 10.1002/bin.301

Arntzen, E., & Hansen, S. (2011). Training structures and the formation of equivalence classes. *European Journal of Behavior Analysis, 12*, 483-503. Retrieved from www.ejoba.org

Arntzen, E., & Haugland, S. (2012). Titration of limited hold to comparison in conditional discrimination training and stimulus equivalence testing. *The Psychological Record, 62*, 243-262.

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

Arntzen, E., & Holth, P. (2000a). Equivalence outcome in individual subjects as a function of training structure. *The Psychological Record, 50*, 603-628.

Arntzen, E., & Holth, P. (2000b). Probability of stimulus equivalence as a function of class size vs. number of classes. *The Psychological Record, 50*, 79-104.

Arntzen, E., & Lian, T. (2010). Trained and derived relations with pictures versus abstract stimuli as nodes. *The Psychological Record, 60*, 659-678. Retrieved from http://opensiuc.lib.siu.edu/tpr/vol60/iss4/8/

Baron, A. (1985). Measurement scales and the age-complexity hypothesis. *Experimental Aging Research, 11*(4), 193-199. doi: 10.1080/03610738508259187

Bentall, R. P., Dickins, D. W., & Fox, S. R. A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology, 46B*, 187-214. doi: 10.1080/14640749308401085

Bentall, R. P., Jones, R. M., & Dickins, D. W. (1998). Errors and response latencies as a function of nodal distance in 5-member equivalence classes. *The Psychological Record, 49*, 93-115. Retrieved from http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1452&context=tpr

Bucklin, B. R., Dickinson, A. M., & Brethower, D. M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13*, 140-163.

Dymond, S., & Rehfeldt, R. A. (2001). Supplemental measures of derived stimulus relations. *Experimental Analysis of Human Behavior Bulletin, 19*, 8-12. Retrieved from http://psy.swan.ac.uk/staff/dymond/pubs/DymondRehfeldt.pdf

Elifsen, C., & Arntzen, E. (2009). On the role of trial types in tests for stimulus equivalence. *European Journal of Behavior Analysis, 10*, 187-202.

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

Fields, L., Landon-Jimenez, D. V., Buffington, D. M., & Adams, B. J. (1995). Maintained nodal-distance effects in equivalence classes. *Journal of the Experimental Analysis of Behavior, 64*, 129-145. doi: 10.1901/jeab.1995.64-129

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*, 317-332. doi: 10.1901/jeab.1987.48-317.

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

Holth, P., & Arntzen, E. (1998a). Stimulus familiarity and the delayed emergence of stimulus equivalence or consistent nonequivalence. *The Psychological Record, 48*, 81-110.

Holth, P., & Arntzen, E. (1998b). Symmetry versus sequentiality related to prior training, sequential dependency of stimuli, and verbal labeling. *The Psychological Record, 48*, 81-110.

Holth, P., & Arntzen, E. (2000). Reaction times and the emergence of class consistent responding: A case for precurrent responding? *The Psychological Record, 50*, 305-337.

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

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*, 617-634.

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

Imam, A. A. (2006). Experimental control of nodality via equal presentations of conditional discriminations in different equivalence protocols under speed and no-speed conditions. *Journal of the Experimental Analysis of Behavior, 85*, 107-124. doi: 10.1901/jeab.2006.58-04

Lyddy, F., Barnes-Holmes, D., & Hampson, P. J. (2000). The effect of stimulus meaningfulness on the formation of equivalence classes. *European Journal of Behavior Analysis,, 1*, 71-87. Retrieved from http://www.ejoba.org/PDF/2000_2/Lyddy_BarnesHolmes_Hampson_2000.pdf

Saunders, R. R., & Green, G. (1999). A discrimination analysis of training structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior, 72*, 117-137. doi: 10.1901/jeab.1999.72-117

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49*(1), 95-115. doi: 10.1901/jeab.1988.49-95

Sidman, M. (1992). Equivalence relations: Some basic considerations. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 15-27). Nevada, Reno: Context Press.

Sidman, M. (1994). *Equivalence relations and behavior: A research history*. Boston: Authors Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*, 127-146. doi: 10.1901/jeab.2000.74-127.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5-22. doi: 10.1901/jeab.1982.37-5.

Spencer, T. J., & Chase, P. N. (1996). Speed analysis and stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65*, 643-659. doi: 10.1901/jeab.1996.65-643

Tomanari, G. Y., Sidman, M., Rubio, A. R., & Dube, W. V. (2006). Equivalence classes with requirements for short response latencies. *Journal of the Experimental Analysis of Behavior, 85*, 349-369. doi: 10.1901/jeab.2006.107-04.

Wang, T., Dack, C., McHugh, L., & Whelan, R. (2011). Preserved nodal number effects under equal reinforcement. *Learning & Behavior : A Psychonomic Society Publication, 39*(3), 224-238. doi: 10.3758/s13420-011-0020-z

Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50*, 125-144. doi: 10.1901/jeab.1988.50-125

Table 1.

Trial Number Completed in Each Training Phase

Speed participants

P9056

Trials	Phase 1[a]	Phase 2	Phase 3
Total	390	108	54
Correct	317	84	52
Incorrect	37	4	0
Missed	36	20	2

P9058

Trials	Phase 1[b]	Phase 2	Phase 3
Total	432	72	1548
Correct	283	58	1206
Incorrect	53	0	11
Missed	96	14	331

P9060

Trials	Phase 1[c]	Phase 2	Phase 3
Total	492	72	420
Correct	319	31	198
Incorrect	97	25	162
Missed	76	16	60

Accuracy participants

P9057

Trials	Phase 1	Phase 2	Phase 3
Total	390	18	54
Correct	356	18	54
Incorrect	34	0	0

P9059

Trials	Phase 1	Phase 2	Phase 3
Total	432	18	54
Correct	414	18	53
Incorrect	18	0	1

P9061

Trials	Phase 1	Phase 2	Phase 3
Total	492	18	72
Correct	461	18	68
Incorrect	31	0	4

Note. Shown is the total number of trials that participants (P9056–P9061) needed to reach criteria in each phase, subdivided into correct, incorrect and missed responses to comparison stimuli. Missed responses could only occur for the speed participants. Phase 1 introduced a titrated LH for the speed participants and a yoked control for the accuracy participants. In Phase 2 the speed participants were exposed to a fixed LH of 1000 ms on comparison stimuli, and the accuracy participants were no longer yoked. In Phase 3 the probabilities of programmed consequences were gradually reduced to 75%, 25% and 0%.

[a] Starting LH value: 4955 ms. [b] Starting LH value: 3927 ms. [c] Starting LH value: 1771 ms.

Table 2

Number of Correct and Incorrect Responses in Testing

	Phase 4[a]					**Phase 5**[b]				
	Correct choices in test					Correct choices in test				
Participants	DT	SYM	EQ	TR	AD	DT	SYM	EQ	TR	AD
9056	11/12	6/12	1/6	1/6	**15/15**	6/12	9/12	3/6	1/6	**15/15**
9057	**12/12**	**12/12**	**6/6**	**6/6**	**15/15**	9/12	10/12	2/6	2/6	**15/15**
9058	**12/12**	**12/12**	**6/6**	**6/6**	**15/15**	**12/12**	**12/12**	**6/6**	**6/6**	**15/15**
9059	**12/12**	11/12	**6/6**	**6/6**	**15/15**	**12/12**	11/12	4/6	5/6	**15/15**
9060	**12/12**	11/12	3/6	5/6	**15/15**	**12/12**	**12/12**	**6/6**	5/6	**15/15**
9061	**12/12**	**12/12**	**6/6**	5/6	14/15	**12/12**	**12/12**	4/6	5/6	**15/15**

Note. Responding that reached accuracy criteria for the directly trained or emergent relations is shown in bold. DT = Directly Trained; SYM = Symmetry; EQ = Equivalence; TR = Transitivity; AD = Adduction; this test was made up of 15 different addition problems, each having a distinct combination of an emergent relation.

[a] First testing. [b] Retention test, which was administered after 14 days, except P9056 and P9060 who returned following 28 and 16 days, respectively.

	1	2
A	ZUL	BAF
B	GIM	ZAY
C	LAL	MIK

Figure 1. Stimuli used for the Pretraining Phase. The numbers 1 and 2 stand for the stimulus sets and the letters A, B, and C the experimenter-defined classes.

	1	2	3
A	ע	כ	ד
B	ה	ל	ב
C	7	8	9

Figure 2. Stimuli used for phases 1–5. The top row of numbers refers to stimulus sets, and the vertically presented letters represent the experimenter-defined classes.

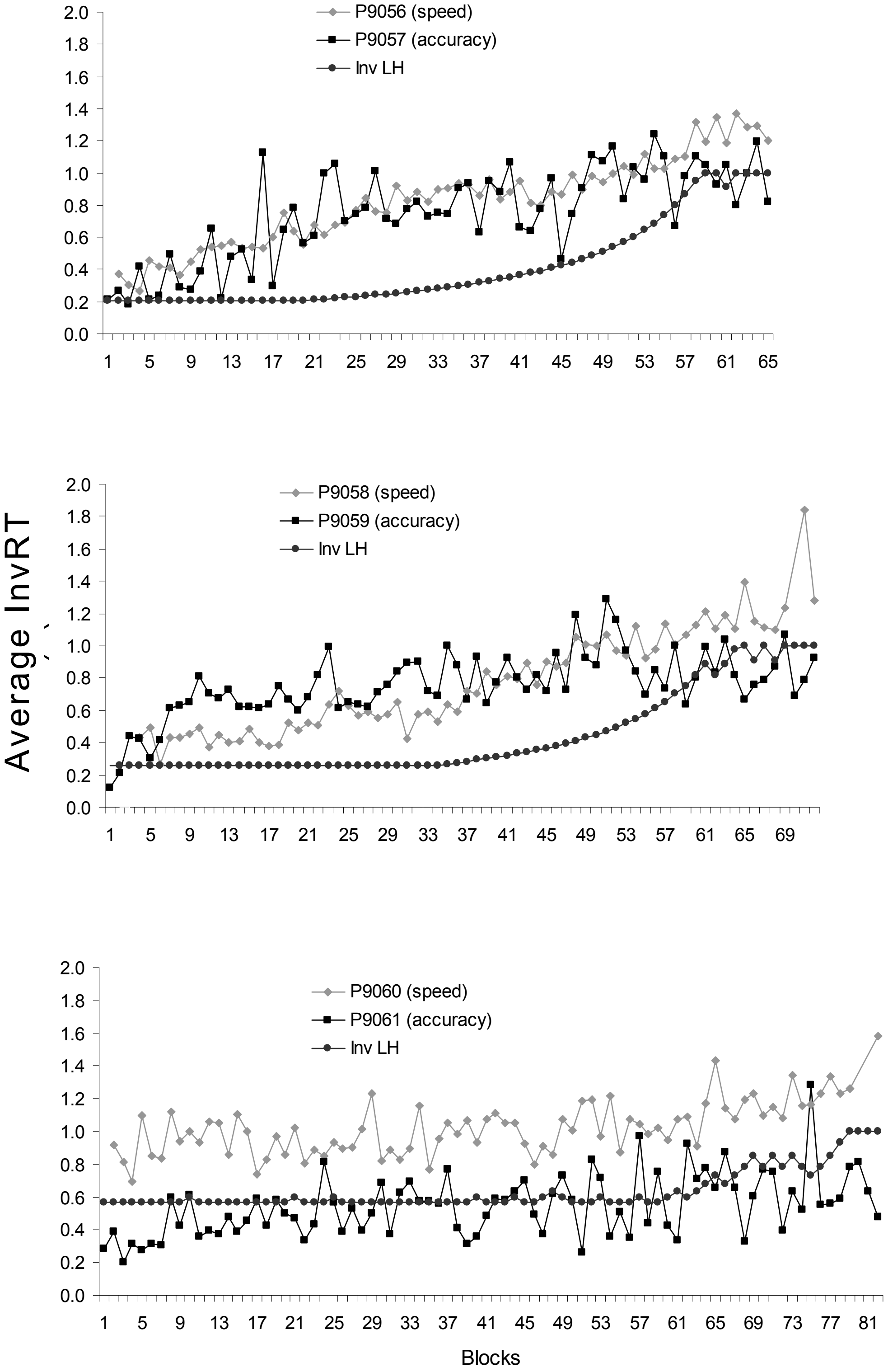

Figure 3. Average Inverse reaction time (InvRT) and the inverse limited hold (InvLH) in each titration block in Phase 1.Graphs are displayed for each matched pair of participants. The average InvRT values from block 1 for P9056 and from blocks 1–3 for P9058 are not displayed, because timed out responses occurred throughout these blocks.

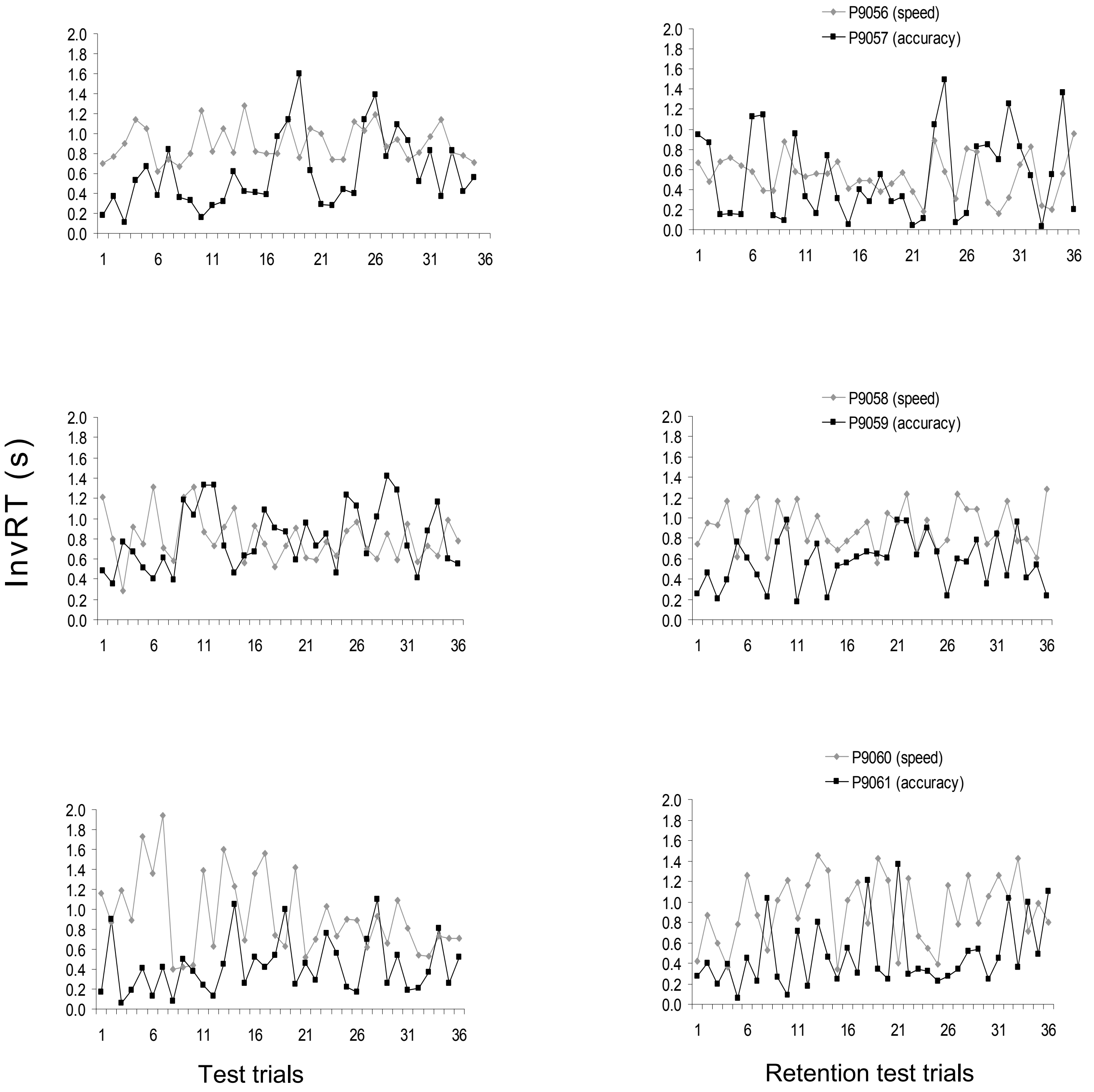

Figure 4. Inverse reaction time (InvRT) in test trials during the first testing and in the two-week retention tests. Individual graphs present the values from matched pairs of speed and accuracy participants.

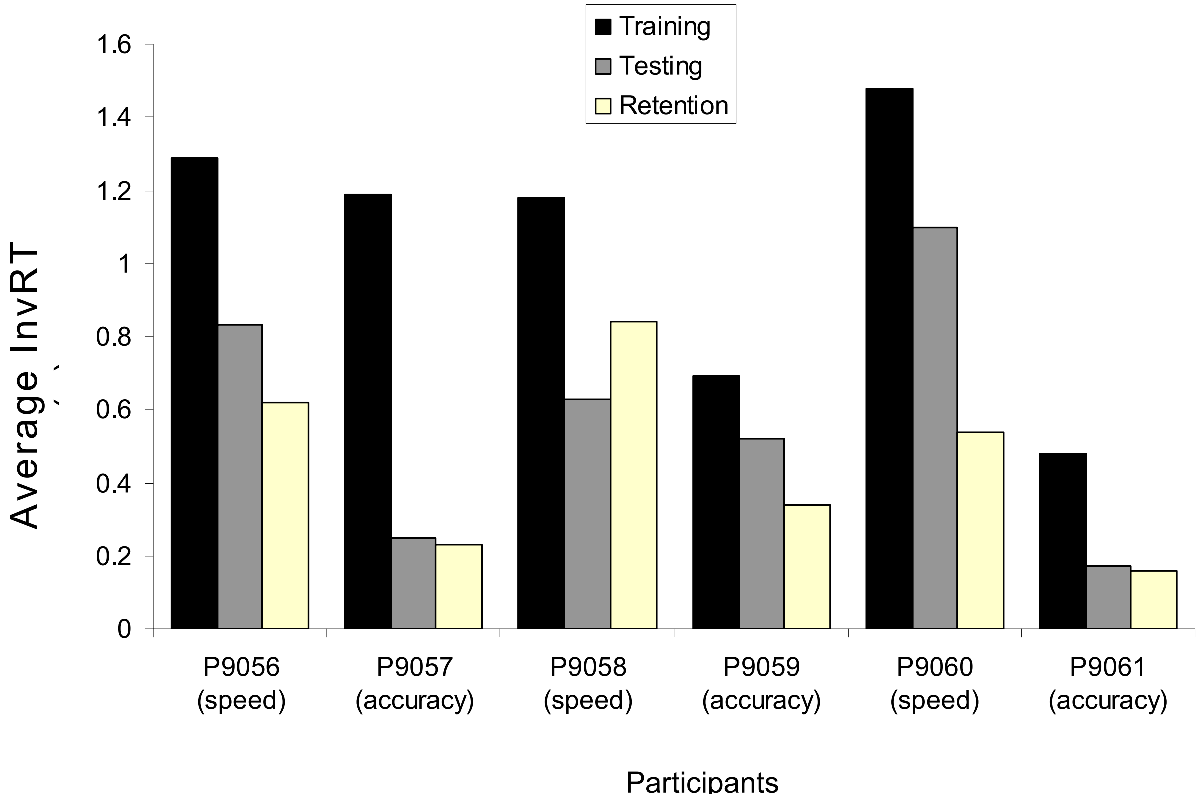

Figure 5. Average inverse reaction time (InvRT) of all participants (P9056–P9061) in the last five trials of training, and the first five trials testing and retention.

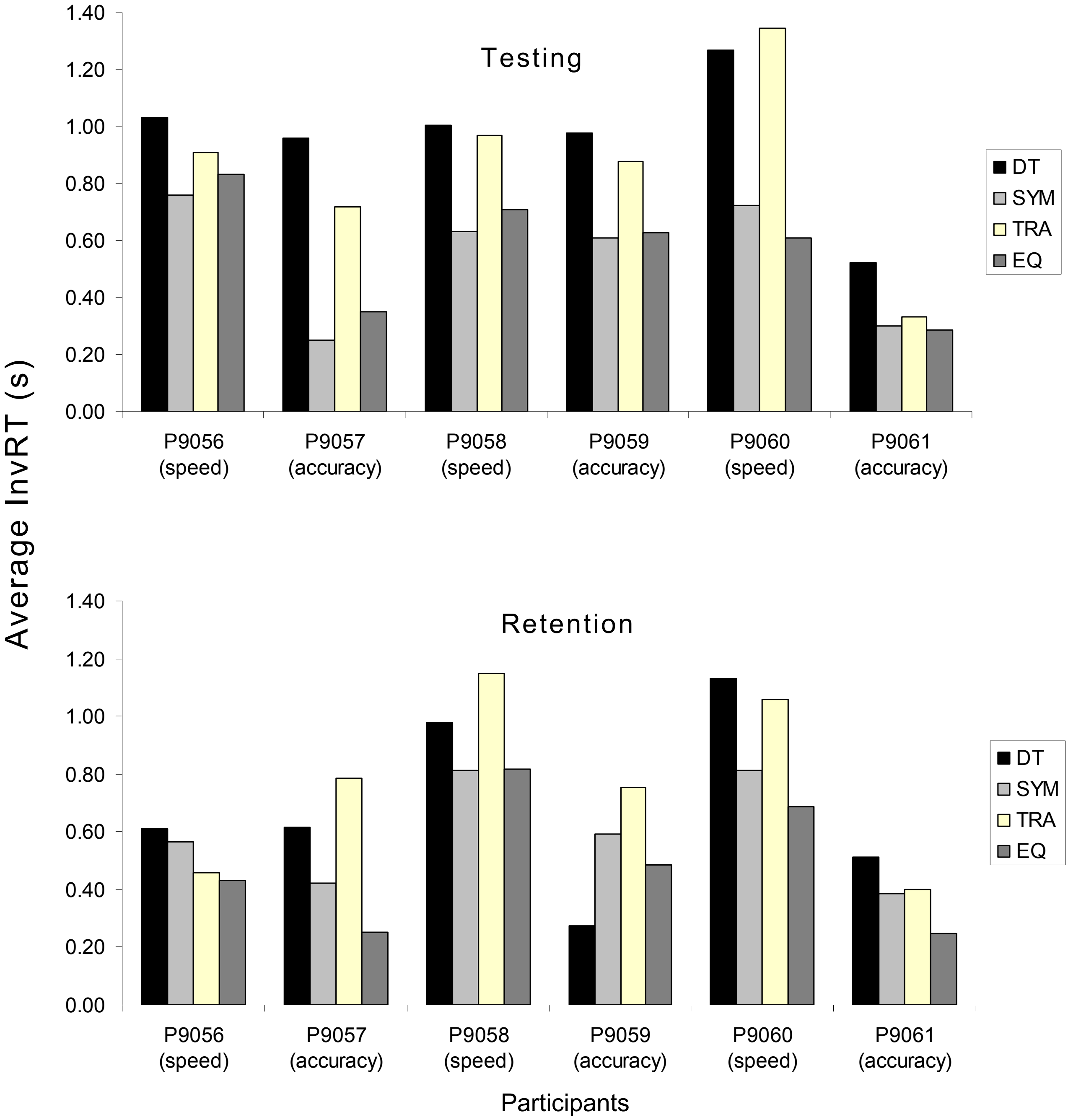

Figure 6. Average inverse reaction time (InvRT) of all participants during the first six trials of directly trained (DT) and symmetry (SYM) trials and all six transitivity (TRA) and equivalence (EQ) trials during testing and retention.

Appendix

ע +ה	ע +ל	ע +ב
כ +ה	כ +ל	כ +ב
ד +ה	ד +ל	ד +ב
ע +כ	ע +ד	כ +ד
ה +ל	ה +ב	ל +ב

Figure 1A. The adduction test adapted from Bucklin, Dickinson & Brethower (2000). Following MTS test trials, the participants were instructed to write the correct answers to the addition problems.